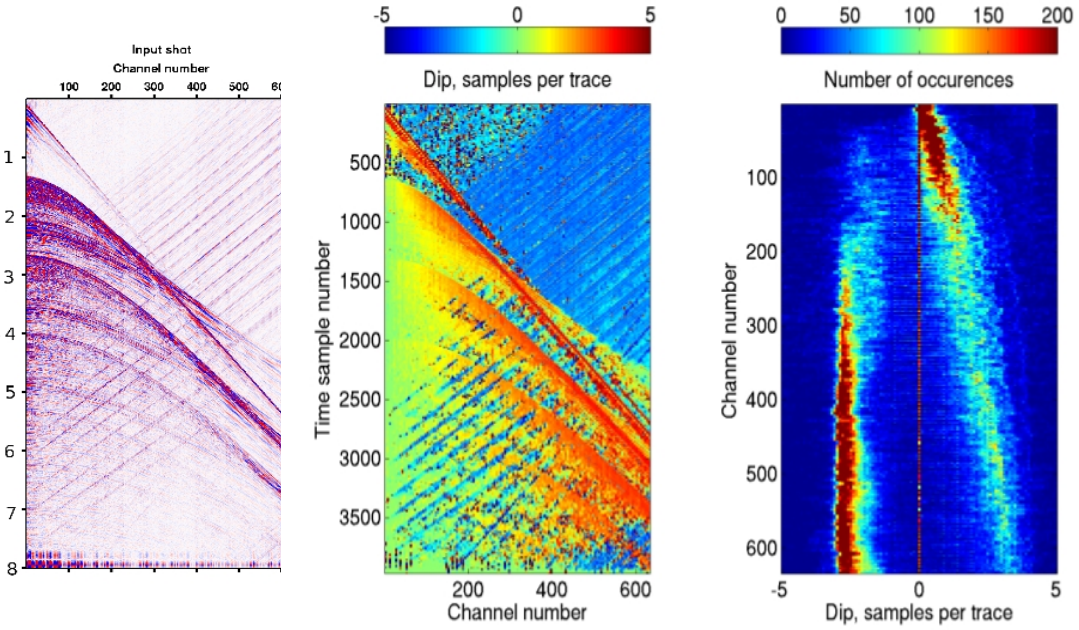
Master Thesis, Department of Geosciences

Identifying Seismic Interference Noise Based on Local Dip Filtering

Milda Grendaite



UNIVERSITY OF OSLO
FACULTY OF MATHEMATICS AND NATURAL SCIENCES

Identifying Seismic Interference Noise Based on Local Dip Filtering

Milda Grendaite

Master Thesis in Geosciences

Discipline: Geophysics

Department of Geosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

June, 2014

Acknowledgments

This thesis was done in collaboration between the University of Oslo and CGG. Most of the work was done at CGG's office in Oslo. First, of all I am thankful for the people who contributed to the developement of this work, especially for my supervisors, Dr. Thomas Elboth at CGG, whose ideas, enthusiasm and daily discussions helped a lot to develop the work, and Prof. Dr. Leiv-J. Gelius at the Department of Geosciences, the University of Oslo, whose constructive comments and care helped to focus and improve the work. I thank Prof. Dr. Valérie Maupin at the Department of Geosciences, the University of Oslo for her initial contribution. I would also like to thank Carl-Inge Nilsen and Vetle Vinje at CGG, R&D for an interesting discussion and Angélique Berthelot, Oleg Tverdomed and Peter Kashchiev at CGG, Processing Department for their help. Next, I thank CGG for providing the data and a work place at the office. Finally, I thank the family for patience and support, friends and colleagues for many discussions and other fun.

Abstract

The aim of this work was to compare various dip estimation methods and apply the best of them to identify the dip of seismic interference noise (SI) in the shot domain. Four dip estimation methods were considered: correlation based method, structure tensor estimation, linear and nonlinear plane wave destruction methods. Both synthetic and real data were used for the tests. The methods were compared with respect to their accuracy, sensitivity to noise, runtime and the number of different parameters the user needs to set. The nonlinear plane wave destruction method was chosen as the one which best meets the requirements. This method was then applied to estimate the dip of SI noise in one seismic line. The noise was then succesfully removed by the $\tau - p$ - common$-p$ SI attenuation method. Dip based signal and noise separation was also tested, however, it removed too much of the signal.

Contents

1 INTRODUCTION

Motivation and aim

In the seismic industry attention has recently been given to methods for so-called dip estimation. These methods provide information about the local dip in the seismic data, and has been applied in areas such as velocity estimation, fault detection, data interpolation and noise attenuation.

A number of dip estimation methods are presented in the literature. They are based on coherency analysis (Marfurt, 2006; Marfurt, Kirlin, et al., 1998; Gersztenkorn and Marfurt, 1999; Marfurt, Sudhaker, et al., 1999), local structure tensor estimation (Vliet and Verbeek, 1995; Morelatto and Biloti, 2013; Randen et al., 2000), plane wave destructors (Claerbout, 1992; Fomel, 2002; Chen et al., 2013a; Chen et al., 2013b), prediction filters (Claerbout, 1992; Claerbout, 2012), edge detection (Luo, Higgs, et al., 1996; Carter and Lines, 2001) and so on.

However, none of the publications known to us, present an evaluation of the dip estimation methods. Even when more than one method is used, it is usually stated that the dip estimates are ”good” but no numerical comparison is given.

Therefore, the aim of this work is not only to estimate the dip, but also to compare various dip estimation methods. To achieve this aim, several different dip estimation methods were implemented and tested on the same datasets.

An example of application of estimated dip is also given. The best dip estimation method was applied to find the dip of seismic interference (SI) noise in one seismic line. The SI noise was subsequently attenuated by the $\tau - p$ - common$-p$ method SI attenuation method which uses information provided by the dip estimation.

SI noise is a type of noise encountered when more than one seismic survey is carried out close to each other, typically, at the vessel separation of 40 - 100 km. The propagation distance of SI depends on water depth and the sea bottom reflection coefficient. SI degrades data quality and is harmful for some prestack processing steps. Its removal can be a challenge.

However, the phenomenon of the SI noise is not fully explained. Most of the publi-

cations related to this noise deal only with its attenuation.

Outline

This work is composed of eight parts. The introduction gives the motivation and the aim of the work. Section 2 presents the background for the topic of the thesis, including the description of the seismic interference noise characteristics. Sections 3 to 5 present the techniques relevant to the work: detailed description of the dip estimation methods, the evaluation of the dip estimates and, finally, the methods of seismic interference attenuation. Moreover, Appendix A presents the methods for inverse problem solution which are relevant to some dip estimation methods and dip-based signal and noise separation. Section 6 presents the main results of tests of the dip estimation methods and an example of application of the estimated dip: SI attenuation from the provided seismic line. Finally, conclusions are given in Section 7 together with some ideas for future work.

2 BACKGROUND INFORMATION

2.1 Seismic Acquisition and Data Description

The data used in this work were acquired in 2013 in the western part of the North Sea, between Shetland and the Faroe Islands using the BroadSeis acquisition technology. The data quality was good although some seismic interference and swell noise was observed.

Twelve streamers, each of almost 8 km in length and equipped with 638 channels, were used. The group interval was $\Delta x = 12.5$, which gives a central midpoint (CMP) interval of 6.25 m in the inline direction. The separation between streamers was 100 m. With two sources 50 m apart shooting in a flip-flop manner, the CMP spacing is 25 m in the crossline direction. The time sampling interval was $\Delta t = 0.002$ s, so temporal frequencies are not aliased up to 250 Hz (valid for noise-free data). Figure 1 shows a seismic vessel in operation.

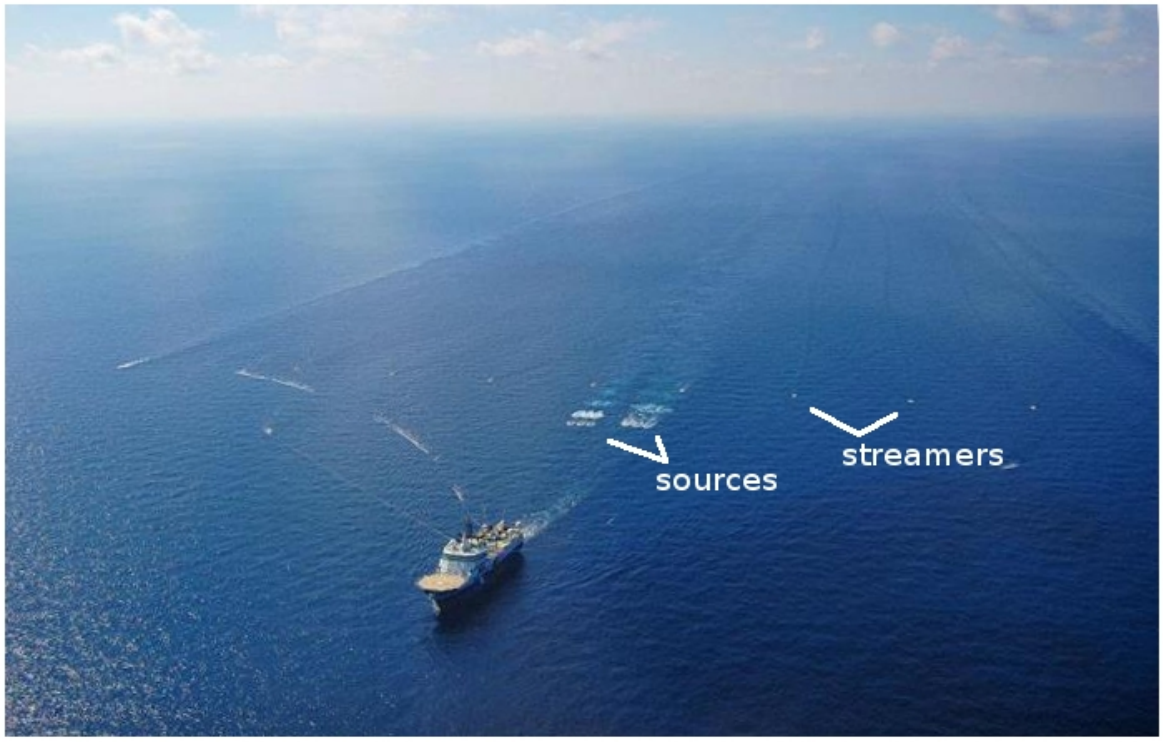


Figure 1: A seismic vessel in operation (*Subsea World News* 2013).

A large number of channels N gives larger fold and an improvement in signal to

noise ratio (SNR $\sim \sqrt{N}$) after stacking and thus helps producing a better subsurface image. However, in case of complex geology a seismic image must be obtained directly from pre-stack migration.

During this acquisition, three component data were recorded: pressure and y- and z- particle motion components. Only pressure data were used in this work because y- and z- components were too noisy (see Section 2.2 for more details).

The data were acquired with streamers slanted in a BroadSeis profile. This means that the streamer's front is at the depth of ~6 m and the tail is at ~50 m. This gives the so called notch (damped frequency) diversity which is beneficial to data processing (Figure 2). For the conventional towing, notches are the same for all receivers: $f_N = v/2z$, where v is velocity of wave propagation in the medium and z is receiver depth. In addition to damped frequencies, some frequencies are boosted (Figure 2). For the slanted towing, notches are different for different recievers because their depth z changes. Boosted frequencies are different as well. Therefore, the overall frequency content does not have notched out frequencies and is broader than the one obtained by the conventional towing (Figure 2).

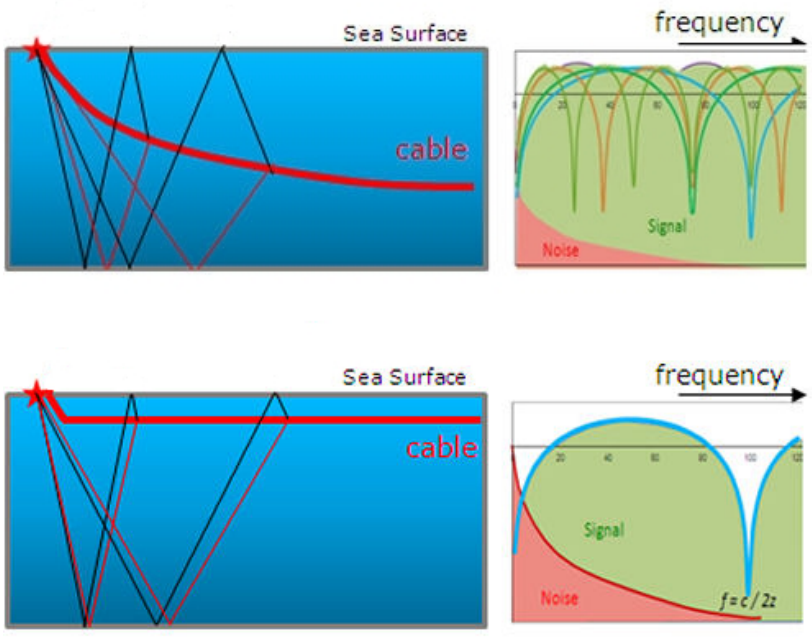


Figure 2: (upper) BroadSeis streamer towing and the frequency band of the record. (lower) For comparison: the same of the conventional towing (*CGG: BroadSeis* 2014).

Typically, a source is an array of airguns deployed at the depth of ∼6 m. It has three source floats with five – six pairs of airguns (airgun clusters) attached to each of the float. A source float is ∼15 m long meaning that airguns are spaced by 3 m while the distance between the two neighboring floats is typically 8 m (*Sercel: Air gun for seismic survey* 2014; Landrø and Amundsen, 2010). This gives the source aperture of ∼15 m. Airguns are of different volumes, thus the pressure pulses they produce are of different duration. Each airgun releases compressed air into water so that the pulse produced by summing the outputs of all the airguns has high amplitude at the begining of the pulse. The subsequent bubble oscillations of each airgun sum up destructively producing a short pulse (Figure 3). Such a sharp pulse resembles a spike and has a fairly flat amplitude spectrum.

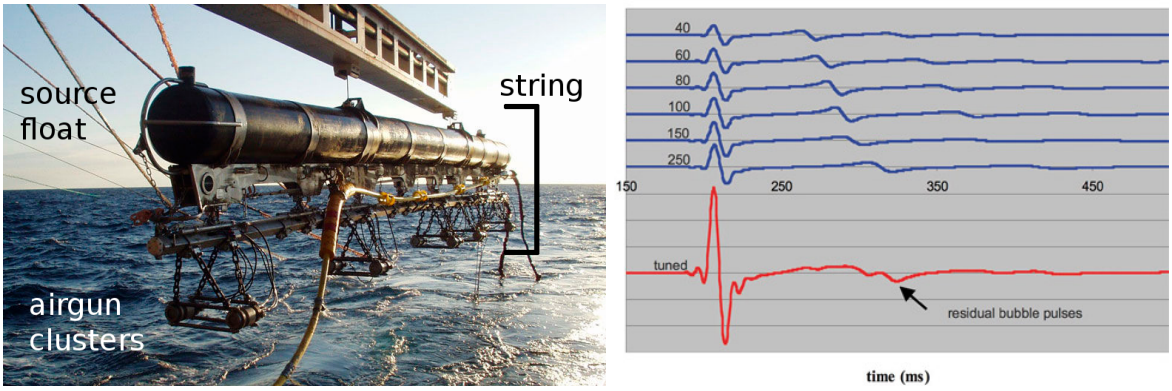


Figure 3: (left) An airgun string (*Sercel: Air gun for seismic survey* 2014). (right) Far-field pressure signatures of individual air-guns vary with gun volume. The summed pressure signature is obtained when the six guns are fired simultaneously. The result is a short pulse of high amplitude (Landrø and Amundsen, 2010).

In this work, a record of one source - cable combination is used.

SI noise is typically removed early in the processing. Therefore, only minor preprocessing was applied to the data: low frequency noise was removed by employing a 3 Hz low-cut filter.

2.2 Types of Noise in Marine Seismic Acquisition

All the undesirable energy seen in seismic recordings is refered as ”noise” while the desired energy is called ”signal” (Scales and Snieder, 1998). The noise originates externally in the recording environment – the ocean, or internally in the

recording system as, for example, vibrations in streamers. The geologic environment can also cause certain types of noise, for example, multiples. However, the latter also carries information about geology.

Ambient noise

The so called Wenz diagram (Wenz, 1962) (Figure 4) shows types of noise present in the oceanic environment. As seen from the diagram, seismic records are affected by most of the noise types existing in the ocean. Weather conditions, surface waves, earthquakes, ocean animals, sea traffic and other contribute to the ambient noise. Even though this figure is taken from an old publication, only the sea traffic noise is expected to have changed – increased.

Vibrations in streamers

There are three types of waves that can propagate in the towed streamers despite that they are engineered to dampen them. These are longitudinal (bulge), transversal and torsional waves (Teigen et al., 2012). Mathematically, transversal waves for a streamer under tension are described by the 4^{th} order partial differential equation of motion, called Euler-Bernoulli equation with dampening:

$$\rho A \frac{\partial^2 u(t,x)}{\partial t^2} = T(x) \frac{\partial^2 u(t,x)}{\partial x^2} - EI \frac{\partial^4 u(t,x)}{\partial x^4} + \alpha \frac{\partial u(t,x)}{\partial t} + f(t,x). \tag{1}$$

Here, t is time and x is space coordinate along the inline axis, ρ is effective density, A is cross-sectional area of streamer, $T(x)$ is tension, I is the moment of inertia, E is Young's modulus (stiffness, stress/strain ratio) of the streamer, α is dampening factor, $f(t,x)$ is external force per unit length. The 4^{th} order term makes the motion dispersive.

The transversal waves in a streamer are low frequency, slow and thus highly aliased waves which severely affect the y- and z- components of the multicomponent surveys (Figure 5, left). These components are affected because they are measured with accelerometers which are sensitive to acceleration by definition. The pressure component sensor – hydrophone – is made acceleration-canceling and, thus, is less sensitive to this type of vibration.

These waves are created by wind induced ocean surface waves passing over streamers. As the waves pass by, the water mass moves circularly and the amplitude of

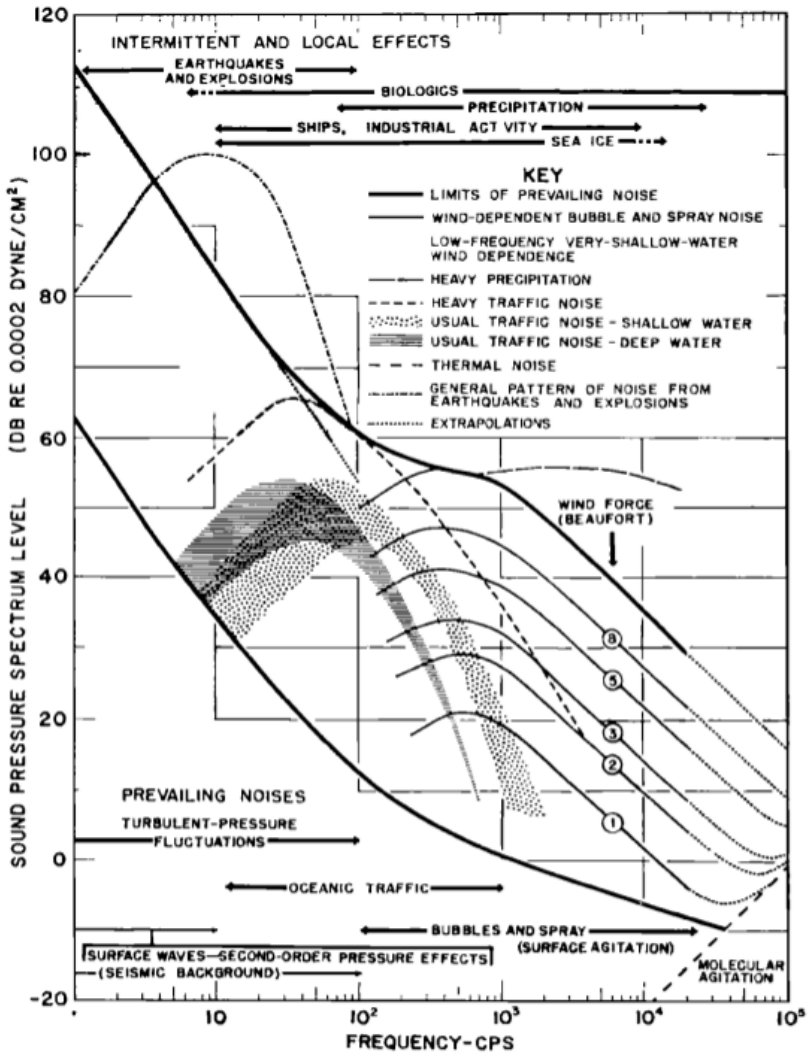


Figure 4: The typical sound levels of ocean background noises versus frequencies (Wenz, 1962).

this motion decreases exponentially with depth (Figure 5, right). Since streamers are towed at the depth where the water mass still moves, the moving water forces them to buckle. This is the case especially close to the birds because birds have larger surface area than a streamer. In pressure records, this manifests as low frequency ($1 - 15$ Hz) noise called swell noise (Figure 6). It can be seen that swell noise affects groups of traces and the number of traces in such groups increase with time because the bird induced vibration/buckling propagates along streamers slowly. Water flow is usually turbulent. This turbulent flow interacts with the streamer surface and creates eddies arround it which also contribute to swell noise (Elboth and Hermansen, 2009).

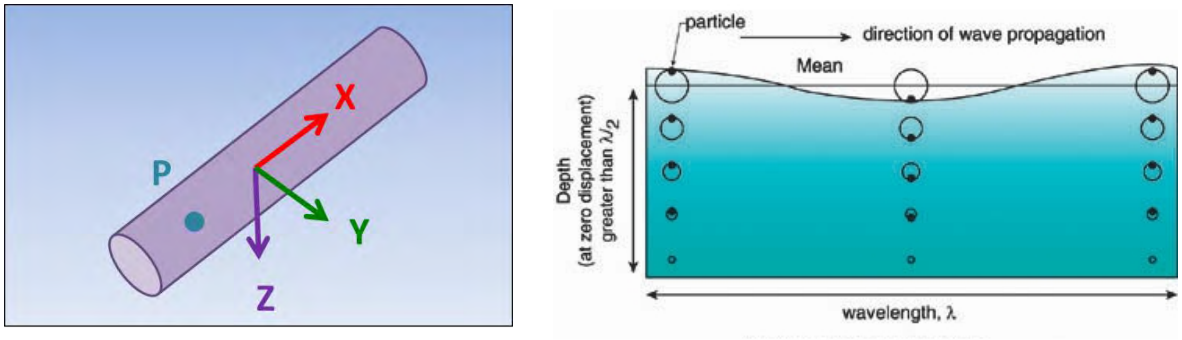


Figure 5: (left) Conceptual view of the multicomponent streamer. P - pressure recorded with hydrophone, X, Y and Z - particle motion components recorded with accelerometers (Teigen et al., 2012). (right) Motion of water mass as wind induced ocean surface waves propagate.

In addition to the swell noise, wind induced ocean surface waves create low frequency (up to 2 Hz) hydrostatic pressure variation noise because the water column heigth above the streamer changes (Figure 7, left). These frequencies are below the normal seismic frequency range and can be removed with a low-cut filter.

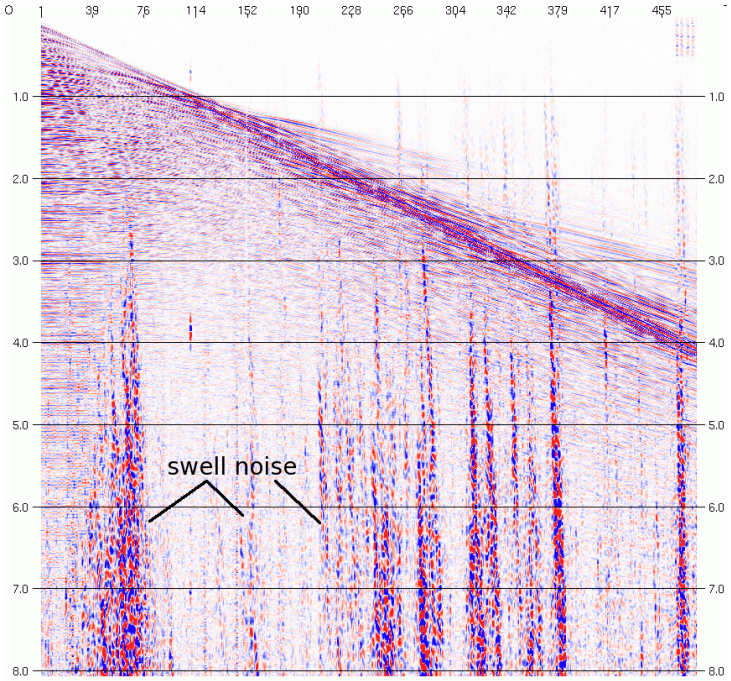


Figure 6: Swell noise (Data Courtesy CGG, R&D).

The longitudinal waves are compressional waves that propagate along the streamer

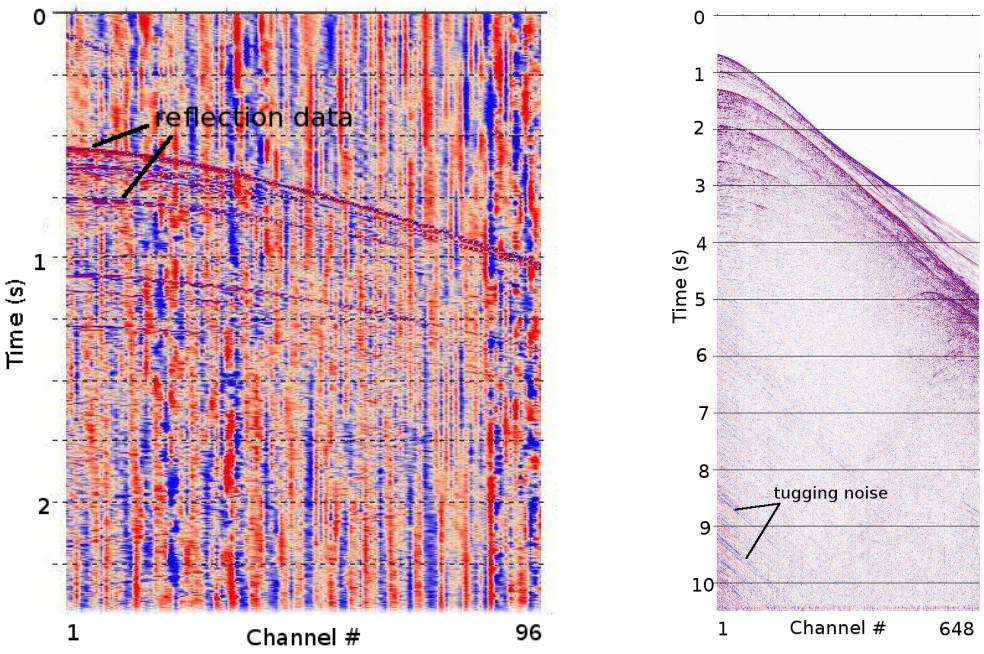


Figure 7: (left) Hydrostatic pressure variation noise. (right) Tugging noise from the vessel (Data Courtesy CGG, R&D).

(in the x- direction). They are described by a simple wave equation:

$$m\frac{\partial^2 u(t,x)}{\partial t^2} = EA\frac{\partial^2 u(t,x)}{\partial x^2} + f(t,x), \tag{2}$$

where m is mass of unit length. These waves affect the x- component of the multicomponent survey (Figure 5). They are more prominent in water filled streamers. They can also be seen in the data acquired with foam filled streamers if the weather conditions were not optimal during the acquisition. This type of vibration manifests as coherent low frequency ($3 - 10$ Hz) noise, called tugging noise, from the vessel and tail buoy (Figure 7, right).

The torsional waves (streamer rotation around its axis) are described by

$$\rho\frac{\partial^2 \phi(t,x)}{\partial t^2} = G\frac{\partial^2 \phi(t,x)}{\partial x^2} + f(t,x), \tag{3}$$

where $\phi(t,x)$ is rotational displacement and G is shear modulus. This type of motion is only minor but might be present.

Streamers are not isotropic, the modula in Equations (1), (2) and (3) are different

for different materials which the streamers are made from. Therefore, they can support waves propagating at different velocities.

2.3 Seismic Interference Noise and Its Characteristics

This detailed section about SI noise is included because a part of this thesis aims to identify the dip of SI noise. Therefore, a good understanding is helpful.

Marine seismic interference noise is a type of coherent seismic noise encountered when more than one seismic survey is carried out in close proximity. The interfering energy mostly arrives as shallow refracted waves and waves reflected in a constructive manner in the water layer (Calvert et al., 1984). If the amplitude of noise exceeds certain limits which are set to assure data quality, seismic vessels usually have to commence on time-sharing which is costly. Seismic vessels have been known to share time since eighties (Akbulut et al., 1985) and as far away as 100 km (Lie, 1988). From the experience, SI is problematic when the distance is less than 40 km (pers. comm. with processors in CGG).

SI appearance in shot gathers

Figure 8 shows five consecutive shot gathers contaminated with SI noise. Here, the noise appears as groups of linear seismic events. Figure 9, left, shows a segment of one channel from the first shot in Figure 8. The SI noise appears as amplitude peaks which might be multiple reflections in water column. Its frequency content is shown in Figure 9, right, and discussed later in this section.

Figure 10, left, shows two vessels acquiring in the same area. The seismic signal generated by the source of vessel B is recorded as SI noise by the streamers of vessel A. Depending on the distance from the source of SI noise to the recording streamer and its direction, the SI noise will have different appearance in a shot gather. Figure 10, right, shows a few SI noise events from different sources. If the SI noise comes from the side and from relatively far away, it reaches the recording streamer as roughly plane waves with dip close to zero, as shown by C1. As the noise source approaches the recording streamer, the registered SI becomes more curved, like C2 and C3 and its dip depends on the channel number. The SI noise appears as plane waves (A or B) with the maximum possible dip, if it comes from

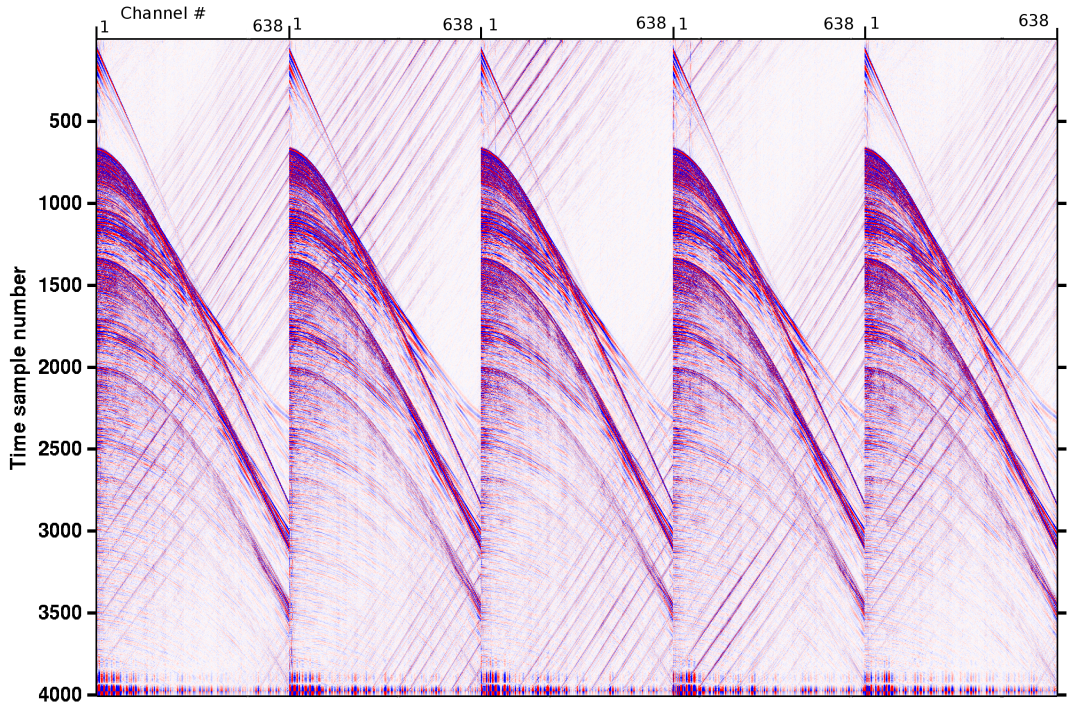


Figure 8: Five consecutive shot records from one seismic line. Only the low cut fiter of 3 Hz was applied. The arrival time of the SI noise train appears random from shot to shot (Data Courtesy of CGG, R&D).

the inline direction. Section 2.4 discusses the possible maximum dip values in detail.

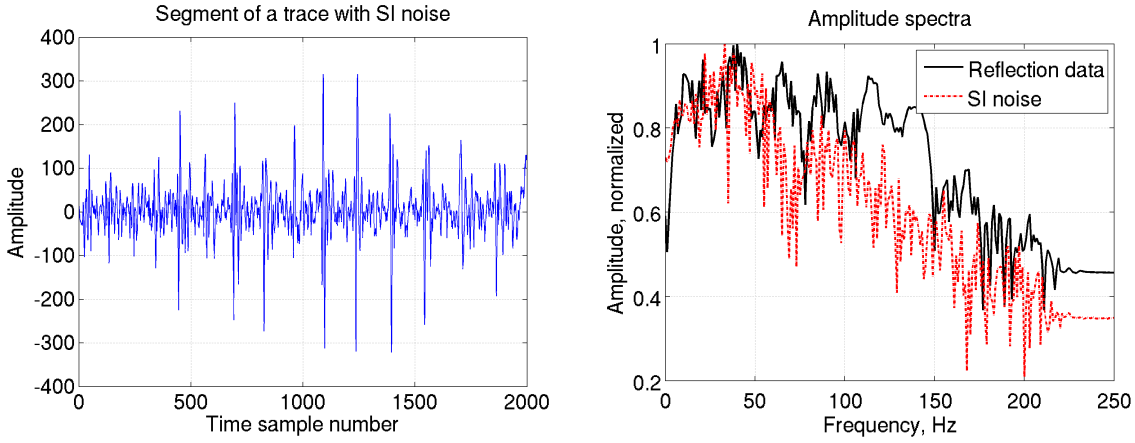


Figure 9: (left) Segment of the channel number 500 from the first shot in Figure 8. (right) Frequency content of the seismic data and the SI noise (the SI noise segment is shown on the left of this figure). Only the low cut filter of 3 Hz was applied.

If we can estimate the dip of the SI events, then we can estimate the direction

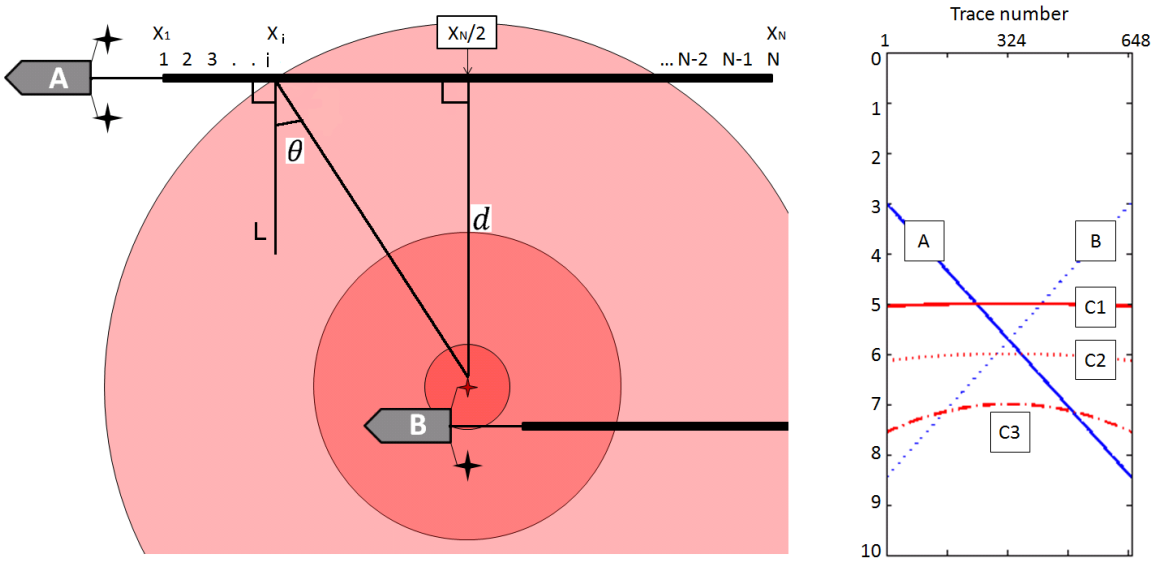


Figure 10: (left) Seismic interference noise arriving at hydrophone x_i of vessel A, originated from the source of vessel B, modified after (Jansen, 2013). (right) SI orginated from: A - inline (front), B - inline (back), C1 - from the side at the distance $d = 100$ km, C2 - from the side at $d = 40$ km and C3 - from the side at $d = 10$ km (Jansen, 2013).

from which the noise came (also see Section 2.4). If the dip of SI depends on the channel, then we may be able to also estimate the distance to the SI noise source.

Periodicity and frequency content

Most often, the arrival times of SI noise appear random from shot to shot (Figure 8) because the sources of the acquiring vessel and the vessel causing noise are not synchronized. Synchronization is not likely because both vessels move at slightly different speeds and in different directions, the firing intervals of both sources and the recording interval of the acquiring vessel are different. Some of the SI attenuation methods (Section 5) rely on the assumption that the SI arrival time is random. If the arrival time is not random and dips are the same as that of the reflection seismic data, many SI attenuation methods will fail.

Frequency filtering is not an appropriate technique for SI noise removal because SI noise is generated by a seismic source which is optimized for the expected frequency band of the seismic data (2 - 150 Hz). Figure 9, right, shows the frequency content of SI noise and of reflection data. In general, their frequency contents highly overlap. The SI noise often lacks high frequencies (higher than 60 Hz in Figure 9, right) because the noise has traveled over a long distance and high frequencies were attenuated.

2.4 Definition of Dip

The definition

Figure 11, left shows a wavefront hitting a streamer. The wavefront hits the point x_{i+1} on the streamer at the time t_{i-1} and the point x_i at the time t_i . The actual distance along the indicent ray traveled in the time interval $\Delta t_i = t_i - t_{i-1}$ is $v\Delta t_i = \Delta x_i^h$ while the distance traveled along the streamer is $\Delta x_i = x_i - x_{i+1}$. Here, v is the wave propagation velocity in sea water, $\sim$1500 m/s. Distance is measured in meters (m) and time in seconds (s). In this situation, Δx_i is negative. The angle θ is the incidence angle of the wavefront's wave vector $\vec{k}$. From geometric relations in Figure 11 we have

$$\sin\theta = \frac{v\Delta t_i}{\Delta x_i} = \frac{\Delta x_i^h}{\Delta x_i}. \tag{4}$$

After rearranging Equation (4) we obtain the expression for the ray parameter or slowness p:

$$\frac{\Delta t_i}{\Delta x_i} = \frac{\sin\theta}{v} = \frac{1}{v_x} = p. \tag{5}$$

Here, v_x is the apparent velocity of the wavefront propagation along the streamer.

The corresponding seismic record is also shown in Figure 11, right. Here, the vector $\vec{d}$ represents the dip vector of the seismic data and from its components we can recover dip $p = \Delta t_i/\Delta x_i$. Therefore, having estimated the dip in the data, the angle θ can also be evaluated (Figure 11, left). Through this thesis we will denote the slowness p for dip.

Limiting values

The angle of incidence θ in Equation (4) ranges in the interval $[-\pi/2, \pi/2]$ and thus the ray parameter p varies in the interval $[-1/v, 1/v]$. Table 1 shows the limiting values of incidence angle θ, dip p and the corresponding direction the SI noise comes from. The maximum absolute dip value is $1/v$ when SI comes along the direction parallel to the streamer.

Units of measurement

In Equation (5), the dip p is measured in seconds per meter, $(\text{m/s})^{-1}$. Dip can also

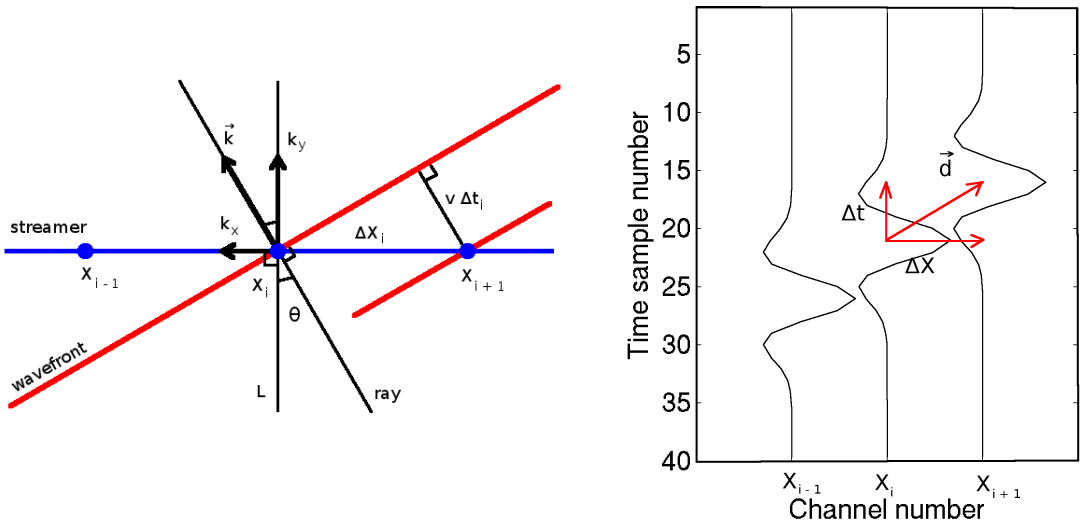


Figure 11: (left) A wavefront hits the point x_i on the streamer at the time t_i at the angle of incidence θ. It has passed the point on the streamer x_{i+1} at time t_{i-1}. The wave vector $\vec{k}$ is perpendicular to the wavefront. (right) The corresponding seismic record. The dip vector $\vec{d}$ and its components along time and space axes Δt_i and Δx_i are shown.

Table 1: Limiting values of the incidence angle at the streamer and the dip of data.

θ	$\sin\theta$	p	Direction
$-\pi/2$	-1	$-1/v$	Inline (back)
0	0	0	From side
$\pi/2$	1	$1/v$	Inline (front)

be expressed in miliseconds per trace, as moveout in miliseconds per shot or in samples per trace:

$$\begin{aligned} p &= \tfrac{\sin\theta}{v}\Delta x_s \cdot 10^3 && \text{ms per trace,} \\ p &= \tfrac{\sin\theta}{v}\Delta x_s n_{chan} \cdot 10^3 && \text{ms per shot,} \\ p &= \tfrac{\sin\theta}{v}\tfrac{\Delta x_s}{\Delta t_s} && \text{samples per trace.} \end{aligned} \tag{6}$$

Here, Δt_s and Δx_s are the temporal and spatial sampling intervals, and n_{chan} is the number of channels in a shot gather. The dip expression in samples per trace is dimensionless. It is particularly useful, because dip estimation methods actually work on data samples. Moreover, it is no longer important, which domain of seismic data we work on.

For the data (shot gathers) used in this work, the temporal and spatial sampling

intervals are $\Delta t_s = 0.002$ s and $\Delta x_s = 12.5$ m. Thus, the absolute maximum expected dip value is equal to 0.0083 seconds per trace or 4.16 samples per trace. In case of $\Delta t_s = 0.004$ s, the maximum possible dip is 2.08 samples per trace.

3 LOCAL DIP ESTIMATION METHODS

Most of local dip estimation methods presented in the literature are based on:

1. coherency measures such as semblance and correlation,
2. covariance matrix estimation,
3. plane wave destructors,
4. and other, such as prediction error filters, complex trace analysis.

All these methods are implemented in the time-space domain. This domain allows accurate local dip estimation. Though dip has an obvious meaning in the frequency domain (apparent velocity $v_x = f/k_x$), the frequency domain does not accurately represent a small subset of discrete data and, thus, is not suitable for local analysis. Figure 12 shows three datasets on successive sparser scales. As the dataset gets sparser, its F-K spectrum becomes blurred.

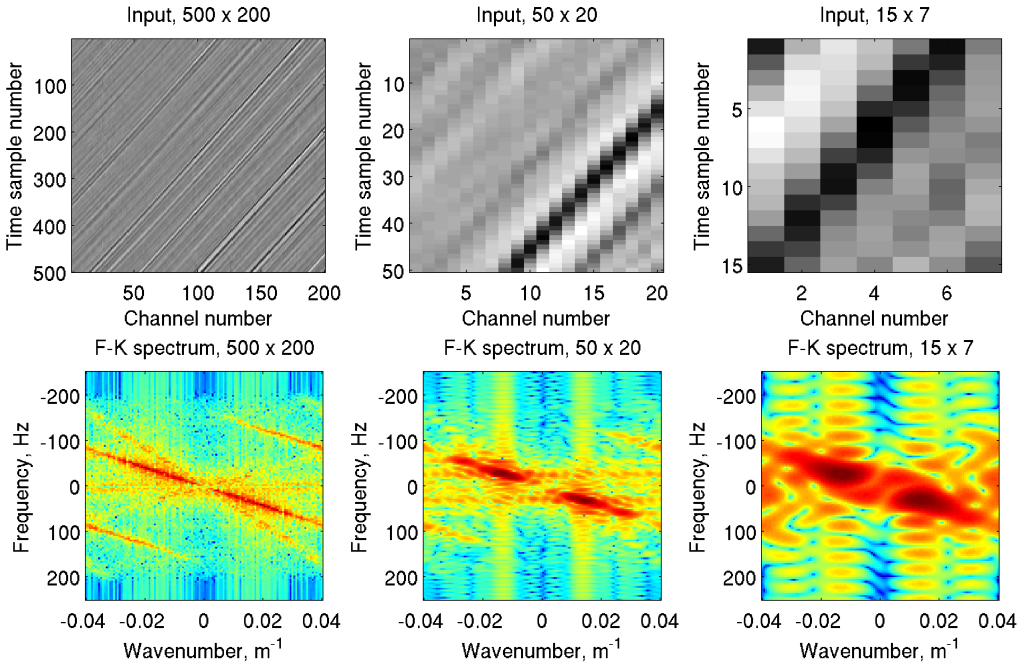


Figure 12: (upper) Input datasets of different size with one constant dip. (lower) The corresponding F-K spectra.

By "local" one means that the dip is estimated within small windows of data. Data are assumed to have one linear event in the window. If there is more than one feature, the methods will estimate the one with the largest amplitude. For example, for the purpose of finding the dip of SI, which is fairly constant, a rather big

window can be extracted from a shot gather where only noise is supposed to be present (e.g., before the first arrivals). For some methods, it is useful to think of two sliding windows. The smaller one is an operator which in this case estimates dip and the larger one is a window in which averaging is done. Averaging may provide a smoother and less noisy result.

The main drawback of the sliding window approach is the choise of the sliding window size. However, its size should reflect the size of the data feature of interest, for example, the wavelet. The window should preferably assign values to its central sample, thus, the window dimensions should be odd numbers of samples.

Based on the underlying theory, dip estimation methods can be divided into deterministic or wave theory based (plane wave destructors), statistical (correlation based method) and hybrid (covariance matrix estimation) methods.

The methods are presented and later tested only for the 2-D case. However, they can be generalized for 3 dimensions by using two perpendicular 2-D filters which estimate the components of the actual dip vector.

In this thesis, most of the methods, such as correlation based method, covariance matrix estimation, local linear plane wave destructor were implemented in MATLAB (version 8.1 (R2013a), The MathWorks, Inc., Natick, Massachusetts, U.S.A.). A few functions from the Madagascar package (version 1.5.3 and version 1.6.3, Madagascar Development Team, Stanford University) were used. In particular, the function ”sfdip” was used for the dip estimation by the nonlinear plane wave destruction method. The function ”sftwodip2” was used to estimate two conflicting dips simultaneously and ”sfplanesignoi” was used for dip-based separation of signal from noise. The dip-based separation of signal from noise is also formulated as an inverse problem (Appendix A).

3.1 Correlation Based Method

The method of cross-correlation measures coherency (similarity) in the seismic data. Here, coherency is a quality measure of two traces being in phase. The correlation method has a lot of applications, for example, automatic horizon pick-

ing (Marfurt, Kirlin, et al., 1998). Moreover, the correlation coefficient itself can be used as an attribute which indicates coherency in the data. The lack of coherency can be applied for detection of faults and stratigraphic features (Bahorich and Farmer, 1995; Carter and Lines, 2001).

There are other methods that measure coherency. They are based on semblance calculation (Neidell and Taner, 1971; Marfurt, Kirlin, et al., 1998), eigendecomposition of the data covariance matrix (Gersztenkorn and Marfurt, 1999; Marfurt, Sudhaker, et al., 1999) and edge detection (Luo, Higgs, et al., 1996; Carter and Lines, 2001). All these methods, except edge detection, are scanning methods, that is, in order to find the dip, all possible dips need to be tested and the one with the largest coherency is selected (Marfurt, Sudhaker, et al., 1999; Marfurt, 2006). Both migrated and stacked seismic data were used in these applications. In this work, we choose the correlation method because it is known to work well and it was successfully applied to nearly unprocessed shot gathers (Jansen, 2013).

Method

The implemented method is based on the normalized cross-correlation. In the sliding window, each two neighboring traces are cross-correlated for the allowed lag values. The coefficient of normalized cross-correlation of two traces x_j and x_{j+1} for lag τ is

$$R_{x_j,x_{j+1}}(\tau) = \begin{cases} \frac{1}{N}\sum_{i=0}^{N-\tau-1} \frac{(A_{i+\tau,x_j}-\mu_{x_j})(A^*_{i,x_{j+1}}-\mu_{x_{j+1}})}{\sigma_{x_j}\sigma_{x_{j+1}}}, & \tau \geq 0 \\ R^*_{x_j,x_{j+1}}(-\tau), & \tau < 0 \end{cases} \tag{7}$$

where N is the number of time samples in the sliding window, asterix denotes the complex conjugate, $A_{i+\tau,x_j}$ and $A_{i,x_{j+1}}$ are amplitudes of the time samples, μ_{x_j} and $\mu_{x_{j+1}}$ are the mean values and σ_{x_j} and $\sigma_{x_{j+1}}$ standard deviations of the traces x_j and x_{j+1} in that window. Lag τ varies from $-N+1$ to $N-1$ but can be limited (MathWorks Inc. Documentation center, 2014).

The lag with the largest absolute value of correlation coefficient represents the vertical (along time axis) component of the local dip vector:

$$d_t = lag(arg(max|R_{x_j,x_{j+1}}(\tau)|)). \tag{8}$$

Here, dip is measured in samples per trace. The dip component along the spatial axis is always set to one because we only consider two neighboring traces.

In order to get a smoother result or when the sliding window contains more than two traces, the mean or median values of lags and correlation coefficients in the window are calculated. The median is prefered because outlying values do not affect it.

Dip can take values other than integer number of the time sample interval, thus, data are too coarsely sampled. Therefore, data will often need interpolation in time. After interpolation, there are $(N-1)f_{int}+1$ time samples instead of initially N within the sliding window. Here, f_{int} is the interpolation factor, typically with values being $10-100$. Interpolation by cubic splines was used in this implementation.

Algorithm
To obtain reliable estimates, the number of time samples in the window needs to be rather large in comparison to other methods. The following steps are performed within the sliding window:

1. interpolate traces in time,
2. correlate each two neighboring traces in the window,
3. find the maximum correlation coefficient values and their lags,
4. calculate median value of lag (dip) and correlation coefficients in that window.

Neighboring windows overlap, so redundant interpolation and correlation is performed as the window slides. To avoid this, traces can first be interpolated in time, then each two neighboring traces in the temporal window can be correlated. The remaining steps are performed with the sliding window.

3.2 Covariance Matrix Estimation

Local dip vectors are estimated from local gradient covariance matrices. In the literature, the gradient covariance matrices are also called gradient-square tensors or structure tensors. Note, that the covariance matrix is an estimate of the structure tensor (Vliet and Verbeek, 1995). Apart from dip estimation, strucure

tensors can be used to estimate coherency and texture of data, detect edges, or filter data along the local dip (Årre, 2013; Vliet and Verbeek, 1995; Morelatto and Biloti, 2013; Randen et al., 2000; Berthelot et al., 2011).

Main steps and algorithm

The implemented method has four main steps which are performed in the sliding window:

1. estimate local gradient vectors for each data point in the window,
2. estimate local gradient covariance matrix,
3. eigendecompose the covariance matrices,
4. calculate attributes from eigenvectors and eigenvalues.

To avoid redundant gradient calculation, gradients should be calculated first. Then a sliding window should be slided over the gradients, not the input data.

Gradient estimation

In the simplest way, the data gradient vector is estimated by the first order Taylor approximation, that is, by applying the forward finite difference scheme along the data coordinates:

$$\begin{aligned} A(t+\Delta t, x+\Delta x) - A(t,x) &= \Delta t A_t^{'} + \Delta x A_x^{'} \\ A_{j+1}^{t+1} - A_j^t &= \Delta t A_t^{'} + \Delta x A_x^{'}, \end{aligned} \tag{9}$$

where A_t' and A_x' are the partial derivatives of A along t and x directions evaluated at (t, x).

The derivatives can be estimated by convolving data with gradient operators, such as finite difference stencils. For example, the second order central finite difference stencil can be constructed from the outer product of the differencing and smoothing kernels: [-1 0 1]$^T\times$ [1 2 1]. This particular stencil is called the Sobel operator (Sobel and Feldman, 1968). The fourth order stencil is constructed similarly. The forward finite difference stencil is introduced in Section 3.3.

The derivative of the Gaussian function can be used as a gradient operator (Vliet and Verbeek, 1995; Randen et al., 2000) or the data in the window can be smoothed with the Gaussian function before calculating derivatives (Morelatto and Biloti,

2013). This intends to suppress high frequency noise and to apply weights on data samples. Then, derivatives are smoothed again by applying weights in the covariance matrix calculation (Vliet and Verbeek, 1995; Morelatto and Biloti, 2013; Randen et al., 2000). Similar approaches are used in image analysis for edge detection. The derivative of an image is estimated by convolving the image with finite difference operators or Gaussian derivative operator (Szeliski, 2010).

Section 6.2 presents the comparison of these stencils as dip filters.

Structure tensor and covariance matrix

The structure tensor is constructed from the gradient vector components. In the matrix notation we have:

$$\begin{aligned} E_j^t(\mathbf{d}) &= \left(A_{j+1}^{t+1} - A_j^t\right)^2 \approx \left(\Delta t A_t^{'} + \Delta x A_x^{'}\right)^2 \\ &= \mathbf{d}^T \begin{bmatrix} A_t^{'2} & A_t^{'} A_x^{'} \\ A_x^{'} A_t^{'} & A_x^{'2} \end{bmatrix} \mathbf{d} = \mathbf{d}^T \mathbf{T} \mathbf{d}, \end{aligned} \tag{10}$$

where the expression $\mathbf{T}$ is called structure tensor. It represents the outer product of the gradient vector $[A_t^{'}, A_x^{'}]$, or the covariance of the gradient vector components $\text{cov}(A_t^{'}, A_x^{'})$. Here, the tensor is calculated from only one sample and, thus, is of rank one. If the energy E_{ij} of this tensor is zero, then the vector $\mathbf{d} = [\Delta t, \Delta x]$ is aligned with the main feature in the data.

Even though the dip of the local main feature of the data is expected to change smoothly, the data contains noise. In order to obtain a robust estimate, weighted neighboring strucure tensors can be included in the estimate of the local structure tensor:

$$E_W(\mathbf{d}) = \sum_{k \in W} \mathbf{d}^T w_k \mathbf{T}_k \mathbf{d} = \mathbf{d}^T \bar{\mathbf{T}} \mathbf{d}, \tag{11}$$

where W represents the window and

$$\bar{\mathbf{T}} = \begin{bmatrix} \langle A_t^{'2} \rangle & \langle A_t^{'} A_x^{'} \rangle \\ \langle A_x^{'} A_t^{'} \rangle & \langle A_x^{'2} \rangle \end{bmatrix} \tag{12}$$

is a weighted sum of structure tensors or the smoothed structure tensor. In case of

the element $\langle A_t^{'} A_x^{'} \rangle$ from Equation (12), we write

$$\langle A_t^{'} A_x^{'} \rangle = \frac{1}{N} \sum_{k \in W}^{N} (A_{t,k}^{'} - \bar{A}_t^{'})(A_{x,k}^{'} - \bar{A}_x^{'}). \tag{13}$$

Here $\bar{A}_t^{'}$ and $\bar{A}_t^{'}$ are averages of the partial derivatives within the window W. Now, the covariance matrix represents an estimate of the structure tensor. If $\bar{A}_t^{'} = \bar{A}_t^{'} = 0$, the two quantities are equivalent. The energy $E_W(\mathbf{d})$ of the smoothed structure tensor is minimal when $\mathbf{d}$ is aligned with the main feature in the data window.

Figure 13 summarizes the main elements being discussed.

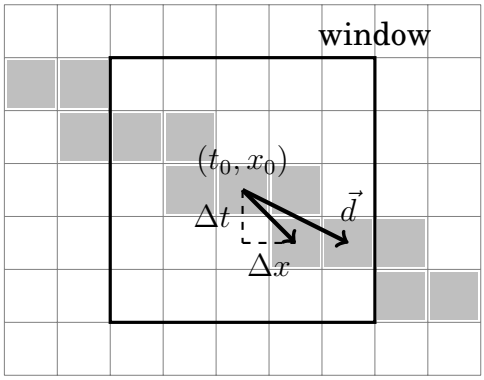


Figure 13: Gradient calculation by forward finite difference scheme for the central sample (t_0, x_0) in a 5×5 window containing a local dominant feature of dip $\vec{d}$.

The weights $\mathbf{W}$ can be of Gaussian function shape, centered in the window and depend on the sample distance from the window center (Morelatto and Biloti, 2013):

$$W_{ij} = \frac{1}{\pi \xi^2} \exp\left(-\frac{i^2 + j^2}{\xi^2}\right). \tag{14}$$

The window size and the standard deviation ξ in Equation (14) are related. Large ξ allows smaller windows, small ξ needs larger ones, so we use a ξ which depends on the window size w: $\xi = -0.25 + 0.25w$. (Luo, Wang, et al., 2006) proposed to use the squared power of complex traces as weights.

Eigendecomposition

The structure tensor has nonegative eigenvalues and orthogonal eigenvectors be-

cause it is positive semidefinite. Eigenvalues are found from the tensor's characteristic equation:

$$(\bar{\mathbf{T}} - \lambda \mathbf{I})\mathbf{v} = 0, \tag{15}$$

where λ is the diagonal matrix of eigenvalues, $\mathbf{I}$ is the identity matrix, and $\mathbf{v}$ is an eigenvector. Since eigenvectors are not zero, the determinant of $(\bar{\mathbf{T}} - \lambda \mathbf{I})$ is equal to zero. Eigenvectors are found by inserting the eigenvalues into Equation (15).
As already being discussed, we have the following criterion $E(\mathbf{d}) \approx 0$ when $\mathbf{d}$ is aligned with the main feature in the data, that is, $\mathbf{d}$ is an estimate of the dip vector. In such a case, the eigenvector $\mathbf{v_2}$ associated with the smaller eigenvalue λ_2 represents the dip vector. The other eigenvector $\mathbf{v_1}$ and eigenvalue λ_1 are asociated with the gradient vector. The dip can be estimated from the components of $\mathbf{v_2}$: v_{2t}/v_{2x}.

Attribute calculation
Other attributes than dip can be constructed to assess the reliability of the dip estimate.

Eigenvalues themselves are attributes because they represent the tensor's energies along the directions of the corresponding eigenvectors. The sum $\lambda_1 + \lambda_2$, where $\lambda_1 \geq \lambda_2$ is the smoothed magnitude (Vliet and Verbeek, 1995). Large magnitudes indicate more reliable estimates.

The ratio of the eigenvalues λ_2/λ_1 is a measure of the absence of one main feature or isotropy of data. It is large when there is no main feature and is small when there is a dominant feature. The opposite attribute represents consistensy of the local orientation or anisotropy: $1 - \lambda_2/\lambda_1$. Large values of anisotropy indicate the presence of one main feature or presumably high signal to noise ratio (SNR) data while low anisotropy shows the lack of main features or presumably low SNR data (Vliet and Verbeek, 1995).

The attribute $(\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$ can be named coherency while the opposing one is called chaos: $1 - (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$ (Randen et al., 2000; Berthelot et al., 2011). Coherency is similar to the anisotropy attribute, chaos to the isotropy attribute.

To summarize, we have these attributes:

sum of eigenvalues	$\lambda_1 + \lambda_2$,
anisotropy	$1 - \lambda_2/\lambda_1$,
isotropy	λ_2/λ_1,
coherency	$(\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$,
chaos	$1 - (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$,
dip	v_{2t}/v_{2x}.

3.3 Plane Wave Destructor

Plane wave model

The idea of plane wave destruction filters was introduced by Claerbout (1992). Locally, a wavefield can be expressed as a superposition of plane waves. In two dimensions (time and space), a wavefront of a plane wave is defined by:

$$u(t, x) = f(t - px), \tag{16}$$

where $u(t, x)$ is the wavefield, f is any arbitraty waveform, t is time, and x is shot $-$ receiver (propagation) distance. The parameter p is called local "dip" or "slope". It can take values in the interval $-1/v \leq p \leq 1/v$, where v is the velocity of wave propagation in the medium.

If data are selected within a small window, all events in that window are assumed to be more or less represented by plane waves. Furthermore, most seismic events can be assumed to be piecewise linear.

Plane wave destructor

A local plane wave can be removed with a partial differential operator (Claerbout, 1992)

$$\left(\frac{\partial}{\partial x} + p\frac{\partial}{\partial t}\right) u(t, x) = v(t, x) \approx 0, \tag{17}$$

where $v(t, x)$ is the residual. Equation (17) represents a differential local plane wave equation and Equation (16) represents its solution.

For discrete data, the derivatives in Equation (17) can be estimated by finite difference stencils. Claerbout (1992) uses the first order forward finite difference stencil with smoothing along the direction, perpendicular to the gradient:

$$\frac{du}{dt} = \frac{1}{2}\left(\frac{u_j^{t+1} - u_j^t}{\Delta t} + \frac{u_{j+1}^{t+1} - u_{j+1}^t}{\Delta t}\right), \tag{18}$$

$$\frac{du}{dx} = \frac{1}{2}\left(\frac{u_{j+1}^t - u_j^t}{\Delta x} + \frac{u_{j+1}^{t+1} - u_j^{t+1}}{\Delta x}\right). \tag{19}$$

Here, the index t refers to t, $t+1$ to $t+\Delta t$ and j to x, $j+1$ to $x+\Delta x$. Time and space increments Δt and Δx are set to $\Delta t = \Delta x = 1$ for convenience.

By using this notation, the differential operator in Equation (17) can be written as

$$\frac{1}{2}\begin{pmatrix} -1-p & 1-p \\ -1+p & 1+p \end{pmatrix}. \tag{20}$$

It works on data samples

$$\begin{matrix} u_j^t & u_{j+1}^t \\ u_j^{t+1} & u_{j+1}^{t+1} \end{matrix} \tag{21}$$

and assigns the estimated dip value to the data sample u_{j+1}^{t+1}. Equation (17) can be written as

$$\mathbf{Au} = \mathbf{v} \approx \mathbf{0} \tag{22}$$

and the operator A is called the plane wave destructor.

Other finite difference stencils or the derivative of the Gaussian function can be used to obtain the data derivative, as mentioned in Section 3.2.

Estimation of the parameter p by linear least squares method

Equation (17) can be rewritten as $\mathbf{x} + p\mathbf{t} = \mathbf{0}$, where x and t are the vectors of data derivatives along x and t axes, respectively. They are computed in a sliding window in which the dip is almost constant.

The parameter p can be found by the linear least squares method (see Appendix A) and is equal to

$$p = -\frac{(\mathbf{x} \cdot \mathbf{t})}{(\mathbf{t} \cdot \mathbf{t})}. \tag{23}$$

The quality measure can be defined as normalized correlation (coherency) (Claerbout, 1992):

$$C = \frac{(\mathbf{x} \cdot \mathbf{t})}{\sqrt{(\mathbf{x} \cdot \mathbf{x})(\mathbf{t} \cdot \mathbf{t})}}. \tag{24}$$

Algorithm

To summarize, the plane wave destruction method can be used to estimate local dip of data and/or to remove the estimated or given dip from the data. Depending on the application, these steps can be carried out in the sliding window:

1. estimate derivatives of the data,
2. find the parameter p by linear least squares method,
3. construct the filter $\mathbf{A}$ (Equation (22)) with the estimated or given dip p and apply it to the data.

Shortcomings

The presented model assumes only one local plane wave (one dip) within the data window. However, it is possible to remove more than one dip by using a different operator in Equation (17).

Finite difference stencils are dispersive for high frequencies (see Section 6.2 for the disccusion), so they fail even before aliasing occurs. Therefore, data needs to be adequately sampled. For undersampled data, the presented method fails to correctly estimate dips larger than one. This is a problem because as discussed in Section 2.4, the maximum possible dip in the data used in this thesis is larger than one.

There are a few methods suggested to overcome this. For example, the dip can also be estimated along the x direction. Then, both estimates are combined into one (Schleicher et al., 2009). Combinations of dip filters can be used to create better dip filters (Hale, 2007). However, it is preferable not to use finite difference stencils. Claerbout (2012) proposes to use cross-correlation when dip is larger than one, or alternatively, a multiscale prediction error filter (Claerbout, 1992; Claerbout, 2012). The best solution is to find an operator with the the best dip filter properties. The next Section 3.4 presents such a solution.

3.4 Improved Plane Wave Destructor

The improved method combines the antialiasing property of prediction error filters with the ability of finite difference stencils to estimate the gradient. The method assumes the same physical model of plane waves as described in Section 3.3. It does not use sliding windows, but estimates dips iteratively by finding a smooth (slowly varying) global solution (Fomel, 2002).

Prediction error filter

If the dip p is time-invariant, the Fourier transform of Equation (17) is an ordinary differential equation:

$$\frac{d\hat{u}}{dx} = i\omega p\hat{u}. \tag{25}$$

with the solution:

$$\hat{u}_x = \hat{u}_0 e^{i\omega px}. \tag{26}$$

Here, $\hat{u}$ is the Fourier transform of the function $u(t, x)$. That is, in the F-X domain, obtaining a trace at x from a neighbouring one at $x - 1$ is equivalent to a multiplication by the time shift operator $e^{i\omega p}$. Equation (26) can now be rewritten as:

$$a_0\hat{u}_x + a_1\hat{u}_{x-1} = 0, \tag{27}$$

where a_0 and a_1 are the coefficients of the F-X prediction error filter: $(a_0, a_1) = (1, -e^{i\omega p})$.

If more than one plane wave is present, they can be predicted by cascading the filters (e.g., applying one after another). Such cascaded filtering can be written in the Z- domain as:

$$A(Z_x) = 1 + a_1 Z_x + a_2 {Z_x}^2 + \ldots = \left(1 - \frac{Z_x}{Z_1}\right)\left(1 - \frac{Z_x}{Z_2}+\right)\ldots. \tag{28}$$

The phase of each solution of the factorized polynomial represents the slope of each plane wave times its frequency: ωp_k, where $k = 1, 2, \ldots$.

Prediction error filtering introduces an error between the actual trace and its predicted value. Such filtering is used to find prediction errors which in turn represent the non-deterministic parts of the data.

Parameter p is a function of time

An important property of local plane waves is that the density of the energy carried by such waves do not change from trace to trace. Therefore, the prediction filter in Equation (26) is an all-pass filter. Equation (26) in Z-transform notation can be written as:

$$\hat{u}_x(Z_t) = \hat{u}_{x-1}(Z_t)\frac{B(1/Z_t)}{B(Z_t)}. \tag{29}$$

Here, Z-transform is defined with $Z = e^{i\omega\Delta t}$ and the temporal sampling interval is set to one, thus, $Z = e^{i\omega}$. The filter $\frac{B(1/Z_t)}{B(Z_t)}$ is the time domain approximation of the phase shift operator $e^{i\omega p}$ (Fomel, 2002; Chen et al., 2013a). This filter is chosen because it has a linear phase response, that is, it preserves the signal shape (Thiran, 1971; Proakis and Manolakis, 1996). Its coefficients can be determined by fitting the filter's response to the response of the phase-shift operator at low frequences, that is, by expanding $e^{i\omega p}$ in a Taylor series at $\omega = 0$. The coefficients of this filter are non-linear functions of p.

Finally, the two dimensional filter in a Z-transform notation is:

$$A(Z_x, Z_t) = 1 - Z_x\frac{B(1/Z_t)}{B(Z_t)}. \tag{30}$$

To avoid deconvolution in Equation (30), a modified filter is used (Fomel, 2002):

$$C(Z_x, Z_t) = B(Z_t) - Z_x B(1/Z_t). \tag{31}$$

A three point filter C can be expressed as the stencil

$$C_3(Z_x, Z_t) = B_3(Z_t) - Z_x B_3(1/Z_t) = \begin{matrix} -b_1 \\ -b_0 \\ -b_{-1} \end{matrix} \begin{matrix} b_{-1} \\ b_0 \\ b_1 \end{matrix} = \begin{matrix} -\frac{(2-p)(1-p)}{12} & \frac{(1+p)(2+p)}{12} \\ -\frac{(2+p)(2-p)}{6} & \frac{(2+p)(2-p)}{6} \\ -\frac{(1+p)(2+p)}{12} & \frac{(2-p)(1-p)}{12} \end{matrix} \tag{32}$$

which acts on the data samples

$$\begin{matrix} u_{j-1}^{t-1} & u_j^{t-1} \\ u_{j-1}^{t} & u_j^{t} \\ u_{j-1}^{t+1} & u_j^{t+1} \end{matrix} \tag{33}$$

Here, indeces t and j stand for time and space coordinates, respectively. The estimated dip p value is assigned to the data sample u_j^t (Fomel, 2002; Chen et al.,

2013a).

Longer filters can be constructed by more accurate Taylor expansion or by deduction. If the numerators of the coefficients of the three point filter are written out like this:

$$\begin{aligned} b_{-1}: &\quad (1+p)(2+p) \\ b_0: &\quad\quad\quad (2+p) \quad (2-p) \\ b_1: &\quad\quad\quad\quad\quad\quad (2-p)(1-p) \end{aligned} \tag{34}$$

and the numerators of the coefficients of the five point filter correspondingly like this:

$$\begin{aligned} b_{-2}: &\quad (1+p)(2+p)(3+p)(4+p) \\ b_{-1}: &\quad\quad\quad (2+p)(3+p)(4+p) \quad (4-p) \\ b_0: &\quad\quad\quad\quad\quad (3+p)(4+p) \quad (4-p)(3-p) \\ b_1: &\quad\quad\quad\quad\quad\quad\quad (4+p) \quad (4-p)(3-p)(2-p) \\ b_2: &\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (4-p)(3-p)(2-p)(1-p) \end{aligned} \tag{35}$$

then longer filters are constructed by deduction. The coefficients in each column add up to 1 (see Equation (32)), so the energy is conserved. The longer the filter, the more accurate it performs at high frequencies (Figure 14).

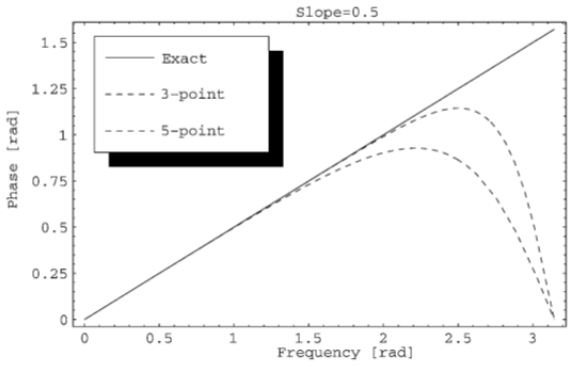


Figure 14: Phase of the time shift operator in Equation (31) and the desired (exact) solution, dip = 0.5 (Fomel, 2002).

Estimation of parameter p as a linear optimization problem

The coefficients of the filters can be found by iterative optimization methods (see Appendix A). Let $\mathbf{C}(p)$ be the operator of convolution in Equation (31). In order to determine the dip, we want to minimize the residual ($\mathbf{d}$ being the data):

$$\mathbf{C}(p)\mathbf{d} \approx \mathbf{0}. \tag{36}$$

The iterative Gauss-Newton method (see Appendix A) is applied where Equation (36) is linearized at each iteration:

$$\mathbf{C}'(p_0)\Delta\mathbf{d} + \mathbf{C}(p_0)\mathbf{d} \approx \mathbf{0}, \tag{37}$$

and is solved for the dip increment Δp. The method requires initial guess p_0.

The system (37) is underdetermined (has more unknowns than equations). To overcome this, additional information is added by the regularization term $\epsilon D\Delta p \approx 0$. Here, D is a gradient operator which enhances small details in the data and makes data less predictable (Fomel, 2002). It is one example of a type of quadratic or Tikhonov regularization term. The parameter ϵ is a scaling constant.

Plane wave destruction filter as an interpolator

Additional insight is presented in (Chen et al., 2013a). A full 2-D Z-transform representation of Equations (29) and (30) is

$$(1 - Z_x Z_t^p)u(t, x) = 0, \tag{38}$$

where Z_t and Z_x are the unit time and unit space shift operators. However, in general, the value at the delay Z_t^p is not sampled. It can be said that it is found by interpolation where the interpolation weights are the coefficients of the filter in Equation (29).

Further improvement of Equation (38) is presented in (Chen et al., 2013b):

$$(1 - Z_x^{p_2} Z_t^{p_1})u(t, x) = 0, \tag{39}$$

where p_1 and p_2 are the dips along t and x axes, respectively. They are constrained by $p_1^2 + p_2^2 = 1$, where $p_1 = \sin\theta$ and $p_2 = \cos\theta$ and with $\tan\theta = p_1/p_2$. In Equation (38) the interpolation is done along the vertical grid line while in Equation (39) it is done on a circle (Figure 15). The filter approximating $Z_x^{p_2}$ is found in a similar way to the the filter in Equation (29). Only the circe-interpolating filter is not aliased (see Section 6.1.3 for the discussion).

Advantages and disadvantages

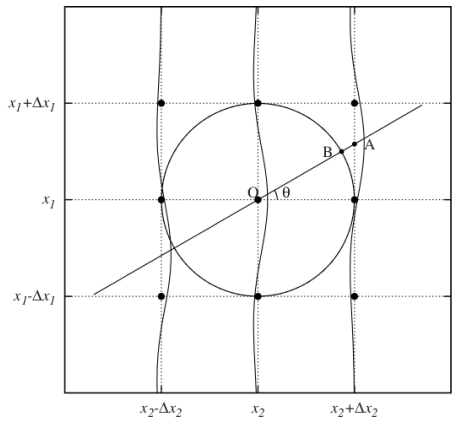


Figure 15: Line-interpolating plane wave destructor interpolates the wavefield along the vertical line at point A while the circle-interpolating plane wave destructor interpolates the wavefield on the circle at point B (Chen et al., 2013b).

The improved method can produce a smooth result (dip field) which is desirable if dips change smoothly in the data.

However, it requires an initial guess of dip values and a good initial guess gives a better final estimate and quicker convergence. Thus, another fast and reliable dip estimation method is probably required to provide a good enough initial guess.

The method is theoretically superior because it has the antiliasing property. Moreover, stencil is small so the method is sensitive to variations in the data.

4 EVALUATION OF THE DIP ESTIMATE

The estimated dips need to be evaluated to assess their reliablity. Moreover, not all the dip estimates of a shot gather represent the SI dip. Statistical methods might be employed to evaluate the reliable dips of SI noise. Since the estimated dip might also depend on the dip estimation method, test datasets were created to assess these methods.

4.1 Test Dataset

Test datasets contain one known constant dip and known signal to noise ratio. The signal part is created in this way: first, one trace is created from random numbers with a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. Then, it is smoothed with a double triangle filter. Finally, it is filtered with a low-cut filter of 5 Hz to ensure that there are no low frequency components. Other traces in the dataset are just shifted copies of this trace. Section 6.1.1 contains examples and tests of such datasets. The noise part was either created from random numbers with Gaussian distribution with mean $\mu = 0$ and standart deviation $\sigma = 1$ or real noise was addded.

The signal to noise ratio in a window of N samples is defined as the ratio of the root-mean-square (RMS) amplitudes of the signal and noise in that window:

$$SNR = \frac{RMS_S}{RMS_N}, \tag{40}$$

where

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} A_i^2} \tag{41}$$

and A_i is the amplitude of sample i.

4.2 Evaluation of Dip Distribution

The output from the dip estimators are dip values at each data sample. A statistical evaluation of the estimated dip values was carried out. If the dip changes in the dataset, the analysis can be done in sliding windows.

Dip distribution in a window usually has one most frequently appearing dip value $\hat{p}$, which we use to represent the actual dip value. There is a certain spread of dip estimates around this value: $\hat{p} \pm \Delta p$. This spread varies depending on the dip estimation method, its parameters and noise (SNR) level in the data. The spread can be characterized by the standard deviation. It is a good approximation of the spread, although in the definition of the standard deviation σ_p

$$\sigma_p^2 = \mathrm{var}(p) = \frac{1}{N}\sum_{i=1}^{N}(p_i - \mu_p)^2, \tag{42}$$

the mean value μ_p is used instead of the most frequent value. Section 6.1.1 contains examples of this type of analysis.

Another way to evaluate the dip estimates is to calculate local standard deviations for each dip sample of the dip output. In this case, the locally most frequent dip value will be close to the local mean value of the dip. The most frequently appearing local standard deviation is chosen to indicate the reliability (noisiness) of the dip estimate. The dip is assumed to change smoothly, so small local standard deviation values indicate the reliable dip estimates. Section 6.1.1 contains examples of such computations.

4.3 Which Dip Values Are Those of SI?

Once the dip values are estimated, we need to identify the dip values which represent SI noise. One way is to calculate the number of occurencies for various dip values and their standard deviations. Since SI noise has rather constant dip, the number of occurencies of these dip values will be large and they will show small standard deviations (Jansen, 2013).

However, in shot gathers, it easier to estimate the dip of the SI noise before the direct and refracted arrivals. The latest recording times of a shot gather are also attractive because the SI amplitude is often higher than that of the signal, but this requires a method which is insensitive to noise.

In general, the SI dip can change significantly within a shot gather if, for example, the noise source is relatively close (see Section 2.3). Then, the traces of the shot gather can be grouped and the SI dip can be estimated for each trace group separately. The SI dip can also change from shot to shot if, for example, the noise source is relatively close and moves relatively to the acquiring vessel.

A procedure which estimates the SI noise in shot gathers can look like this:

For each shot gather repeat:

1. read in a shot gather and prepare it for the dip estimation,
   (a) filter low frequencies,
   (b) apply sensitivity correction - optional,
2. call dip estimation method and apply it on the defined patches in the gather,
3. evaluate the dip estimate based on coherency measure, RMS, standard deviation, local standard deviation and accept only the reliable estimates.

When all the shots are processed, do this:

1. from the accepted estimates create the volume of number of occurences of dip value by accepting the most frequent dip value in each subpatch (group of traces),
2. evaluate the volume of occurences - SI dip should change smoothly if change at all, be localized (do not have outlyers),
3. provide the number of occurences for the SI noise attenuation method ($\tau - p$ - common$-p$) and apply it to obtain SI-attenuated seismic line.

The function of number of occurences of dip values might have gaps if no reliable dip estimate was found. However, SI dips do not change abruptly from shot to shot or from channel to channel, thus, missing values can be interpolated from their neighbors.

5 APPLICATION EXAMPLE: SI REMOVAL

This section shortly overviews the SI attenuation methods. The $\tau - p$ - common$-p$ SI attenuation method is presented in detail because it is used in this thesis (Section 6.4.1). A novel approach of dip-based signal and noise separation is also presented in this section and tested in Section 6.4.2.

5.1 SI Removal Methods

There are two main groups of SI attenuation methods. The first group of methods randomize SI by sorting or transforming the seismic data into the domain where SI is random while the signal remains coherent. For example, SI appears random in common offset (Gulunay, 2008), common midpoint or common$-p$ (Elboth and Hermansen, 2009) domains. The randomized SI noise can be attenuation with random noise attenuation techniques, such as prediction filters (Gulunay et al., 2005) or time-frequency de-noising (Elboth, Presterud, et al., 2010). The approach relies on the assumption that the signal and the SI noise sources are not synchronized and, therefore, the SI arrival time from shot to shot is random.

The other group of methods builds a model of SI noise and subtracts it from the data adaptively. Noise source can produce noise continuously (a passing ship) or in pulses (backscattered energy from big objects). A model of SI can be created by scanning for the apparent position of the noise source. Having estimated the position, the noise arrival times can be predicted for each trace. Then, the noise event is flattened and filtered in the F$-$K or $\tau - p$ domains (Warner et al., 2004; Fookes et al., 2003). If the noise pattern is assumed to be the same from shot to shot, the SI from one shot can be adaptively subtracted in a least squares sense from other shots (Kommedal et al., 2007; Brittan et al., 2008). This approach requires an accurate SI model.

5.2 Tau-p - Common-p Method

This section presents an example where the estimated dips can be applied. The SI noise is attenuated by the $\tau - p$ - common$-p$ method (Elboth and Hermansen,

2009), in which the randomized SI noise is attenuated with a random noise attenuation technique, such as the time-frequency de-noising routine, TFDN, (Elboth, Presterud, et al., 2010). The actual SI dip range $\hat{p} \pm \Delta p$ can be provided by one of the dip estimation methods discussed in Section 3.

The $\tau - p$ - common$-p$ method is comprised of these steps (Figure 16):

1. the shot gathers are transformed into the $\tau - p$ domain,
2. the $\tau - p$ transformed gathers are resorted into common$-p$ domain,
3. a random-noise attenuation technique (such as TFDN) is applied across the SI dip range which output the SI noise,
4. the filtered data (SI) is sorted back into the $\tau - p$ domain
5. and finally transformed back to the shot domain.

This flow produces the SI noise model which is adaptively subtracted from the original data in a least squares sense. The noise output is prefered to the data output because it harms the data less. To summarize, the method combines both approaches of SI removal: it randomizes the SI noise and produces the SI model.

Tau-p transform

The Radon transform is an integral transform of the data along defined curves. There are three commonly used types of curves, that is, line, hyperbola and parabola, which determine the type of the Radon transform. The linear Radon transform is often refered to as the $\tau - p$ transform or slant stack (Diebold and Stoffa, 1981). Summation is done along lines (which in turn represent phase fronts of plane waves) defined by zero-offset intercepts τ and slopes p (Weisstein, 2014b):

$$R(\tau, p)[f(t, x)] = \int_{L=\tau \pm px} f(\tau \pm px, x). \tag{43}$$

The $\tau - p$ transform maps plane-waves into points, so it is the sparse plane-wave representation of seismic data. As mentioned, SI noise often appears as nearly plane-waves in the shot domain, thus, the noise can be localized by mapping into the $\tau - p$ domain. This localization lends itself for muting in the $\tau - p$ domain (Jansen, 2013). However, in general, such muting affects the signal because SI dip can coinside with that of the underlying reflection data.

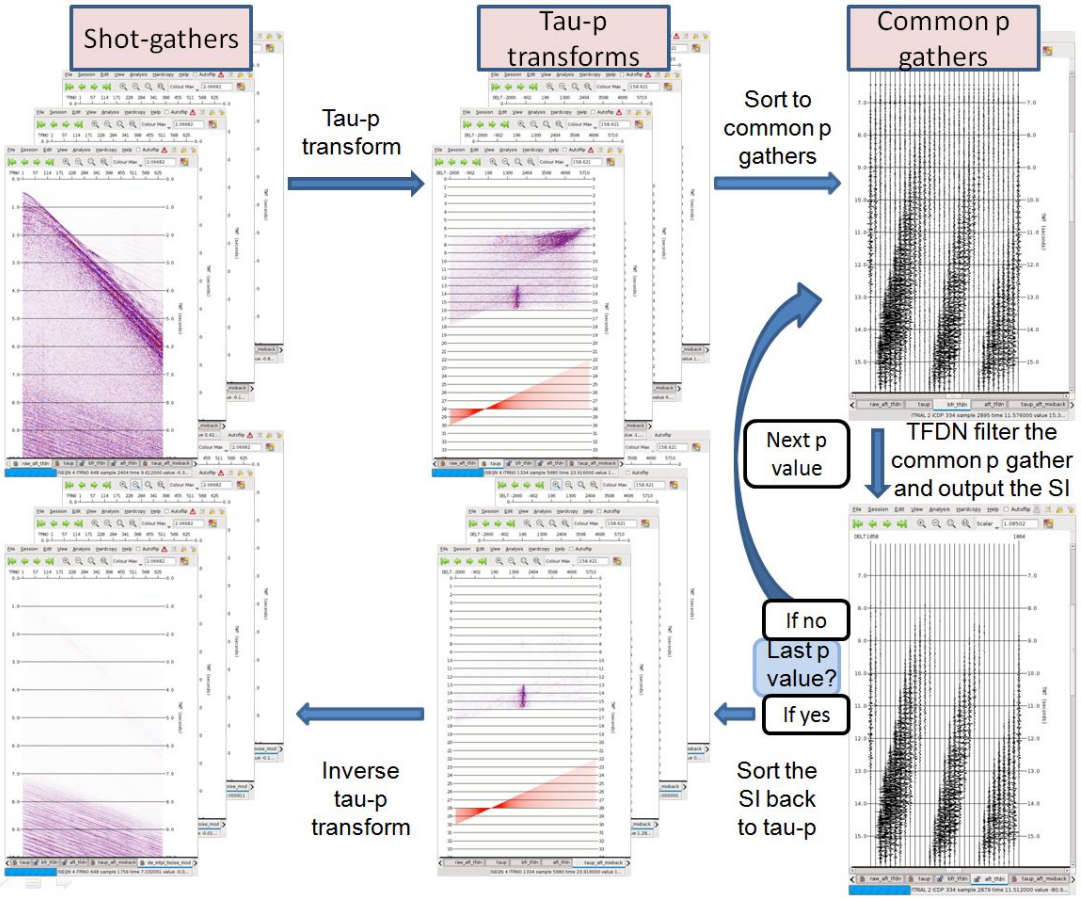


Figure 16: The main steps of the $\tau - p$ - common$-p$ SI noise attenuation method (Jansen, 2013).

A few parameters need to be defined by the processor: the maximum and minimum p values, the p increment, maximum and minimum zero-offset times τ of interest. Figure 17 shows how a 2-D shot gather being mapped to the $\tau - p$ domain. The maximum p is represented by blue color, the minimum p with red color and green color indicates the p of the SI noise.

Common-P sorting

Once all the shot gathers are $\tau - p$ transformed, they can be sorted into common$-p$ (common slowness) domain. If the assumption of non-synchronous sources is satisfied, SI will appear random in this domain. Figure 18 shows an example of such sorting.

Time-frequency de-noising

The time-frequency de-noising technique works in sliding windows. The ampli-

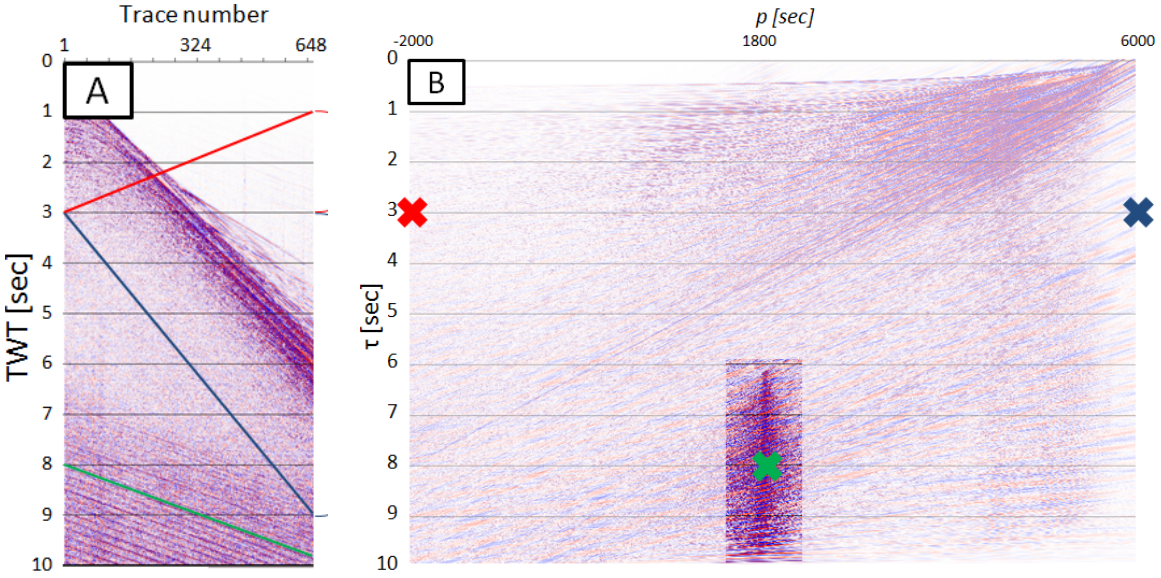


Figure 17: A shot gather with SI noise and its $\tau - p$ transform. Red, blue and green points in the $\tau - p$ domain were obtained by summing the data in the shot domain along the lines of respective colors. The SI noise is localized in the $\tau - p$ domain (Jansen, 2013).

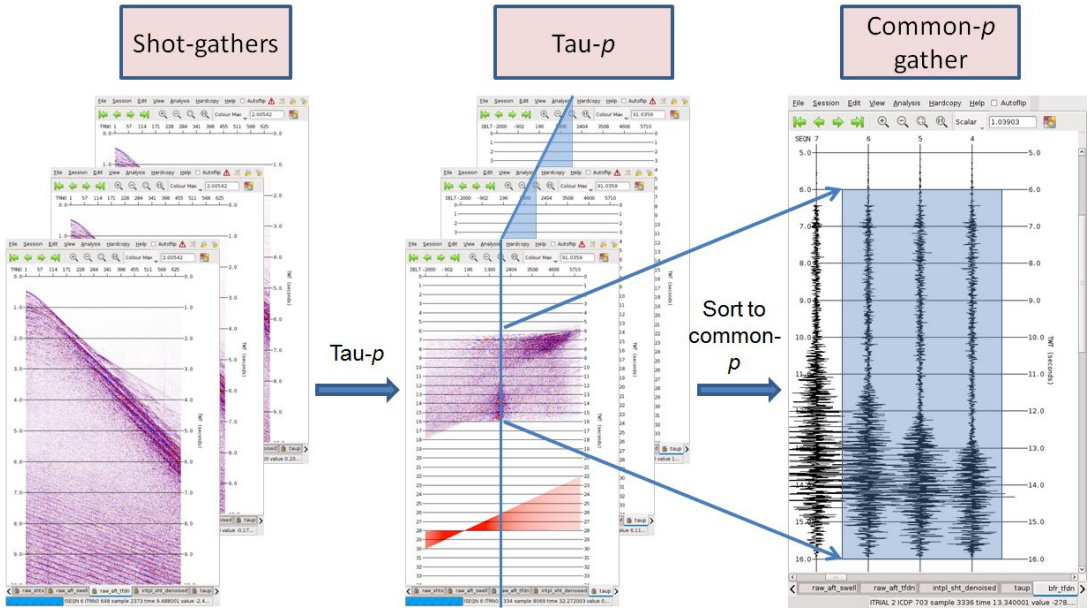


Figure 18: A few shot gathers are transformed into the $\tau - p$ domain and then sorted into the common$-p$ domain (Jansen, 2013).

tude spectrum of the data in the window is estimated. Then, the amplitude at each frequency is compared with the amplitude of the reference trace at that frequency. The abnormally high amplitudes are suppressed. The corrected spectrum is then transformed back into the orginal domain. The successful application of the method depends on the choise of the reference trace and the threshold value for the abnormal amplitudes.

Processing flow

The processing flow applied (Section 6.4.1) to one seismic line, with the aim to illustrate the $\tau - p$ - common$-p$ method, can look like this:

1. raw shot gathers are filtered in order to remove the low frequency components and swell noise,
2. the filtered shots are transformed into $\tau - p$ domain,
3. the $\tau - p$ transformed gathers are sorted into common$-p$ domain,
4. time-frequency de-noising is applied in the common$-p$ domain across the SI dip range, found by a dip estimation method, producing SI-attenuated common$-p$ gathers,
5. the difference between the original common$-p$ sorted data and the SI-attenuated common$-p$ gathers produces the SI noise model in the common$-p$ domain [1],
6. the SI model in the common$-p$ domain is sorted into $\tau - p$ domain,
7. the SI model in the $\tau - p$ domain is transformed back to shot domain and represents the SI model in the shot domain,
8. the difference between the originally filtered shots and the SI model in the shot domain gives the SI-attenuated data in the shot domain.

Therefore, both the SI-attenuated data and the SI noise model are obtained. Correspondingly, two stacks are can be generated and compared:

1. both the original filtered shots and the SI-attenuated shots are sorted to the common midpoint (CMP) domain,
2. CMPs are normal moveout (NMO) corrected,
3. the NMO-corrected CMPs are stacked,
4. the stacks are migrated by Kirchhoff migration,
5. the difference between the original image and the SI-attnuated image is obtained.

5.3 Dip Based Signal and Noise Separation

This section presents how dip can be used to separate signal from noise (Fomel, 2002).

[1] The SI-attenuated common$-p$ gathers can also be sorted into $\tau - p$ domain and transformed into shot domain. However, this does more damage to the signal because $\tau - p$ transform does distorts the signal due to inaccurate amplitude factors in the transform in 2-D data.

The main assumption is that the signal part and the noise part have different dips. This might be the case for ground roll noise or SI noise, both of them being coherent. If we want to remove such noise without randomizing, we need a technique for coherent noise attenuation. The method presented here, first simultaneously estimates two dips: one of the signal and the other of the noise. The two-dip estimator is an extension of the one-dip estimator, presented in Section 3.4. The operator $\mathbf{C}(p)$ in Equation (36) is modified into $\mathbf{C}(p_1)\mathbf{C}(p_2)$ and now represents a cascade of two operators with different dips.

Having estimated both dips, the events are separated based on their dips. This separation is formulated as an inverse problem

$$\begin{cases} \mathbf{C}(p_1)\mathbf{d}_1 \approx \mathbf{0}, \\ \epsilon\mathbf{C}(p_2)\mathbf{d}_2 \approx \mathbf{0}. \end{cases} \tag{44}$$

Here $\mathbf{C}$ is the dip filter, $\mathbf{d}_1$ is the signal component with dip p_1 and $\mathbf{d}_2$ is the noise component with dip p_2. Both parts are constrained by $\mathbf{d}_1 + \mathbf{d}_2 = \mathbf{d}$, where $\mathbf{d}$ is the original input data. The scalar ϵ reflects the SNR of the input data.

The separation is done by solving the equation system (Fomel, 2002)

$$\begin{cases} \mathbf{C}(p_1)\mathbf{d}_2 \approx \mathbf{C}(p_1)\mathbf{d}, \\ \epsilon\mathbf{C}(p_2)\mathbf{d}_2 \approx \mathbf{0}. \end{cases} \tag{45}$$

An example of the application of this technique is presented in Section 6.4.2.

6 RESULTS AND DISCUSSION

6.1 Tests of Dip Estimation Methods

6.1.1 Tests on Synthetic Data

The methods were tested on synthetic datasets of known dip and signal to noise ratio. Two synthetic datasets of size 500 $\times$ 200 (times samples $\times$ number of channels) were generated as described in Section 4. The dip value is set to 0.4 and 3.0 samples per trace, respectively (Figure 19). Two types of noises were added to the clean test datasets: random noise and real noise (Figure 20). The random noise has a Gaussian distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The real noise was taken from a test survey, which was acquired when the vessel was not shooting. It is dominated by the tugging noise from the vessel which appears as plane waves. Thus, it is not random, but fairly coherent. Both types of noises were added to the data with different strengths. The following signal-to-noise (SNR) values were employed: Inf, 100, 50, 20, 10, 5, 2, 1 and 0.5; e.g., the percentages of the noise amplitude with respect to the signal amplitude were 0.0, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0 and 2.0, respectively.

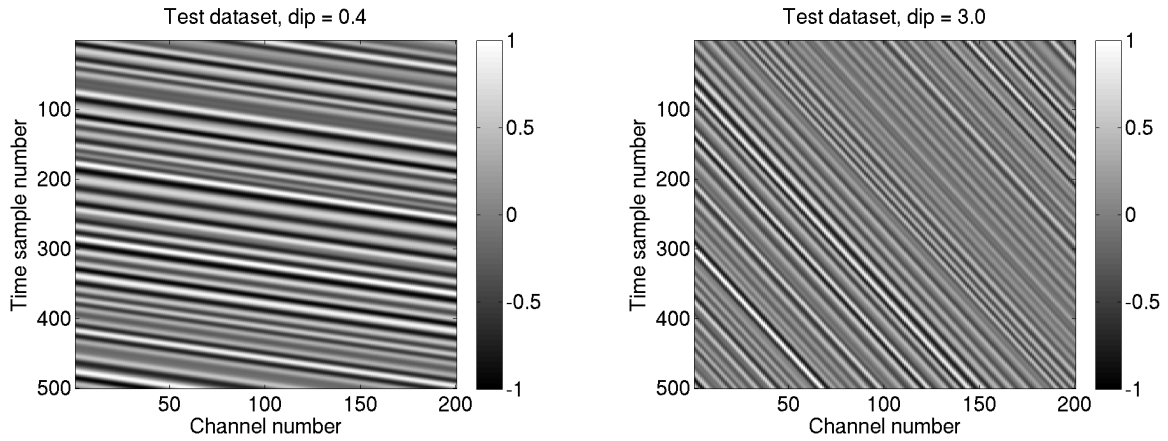


Figure 19: Test datasets. (left) Dip value is 0.4 samples per trace. (right) Dip value is 3.0 samples per trace.

The dip estimates were evaluated in the way presented in Section 4. Three parameters characterize the estimate: the most frequent dip value (actual dip value), the standard deviation of the dip estimate and the most frequent local standard deviation (both reflecting the ”noisiness” of the dip estimate). Ideally, both standard deviations should approach zero because the test datasets contain one constant

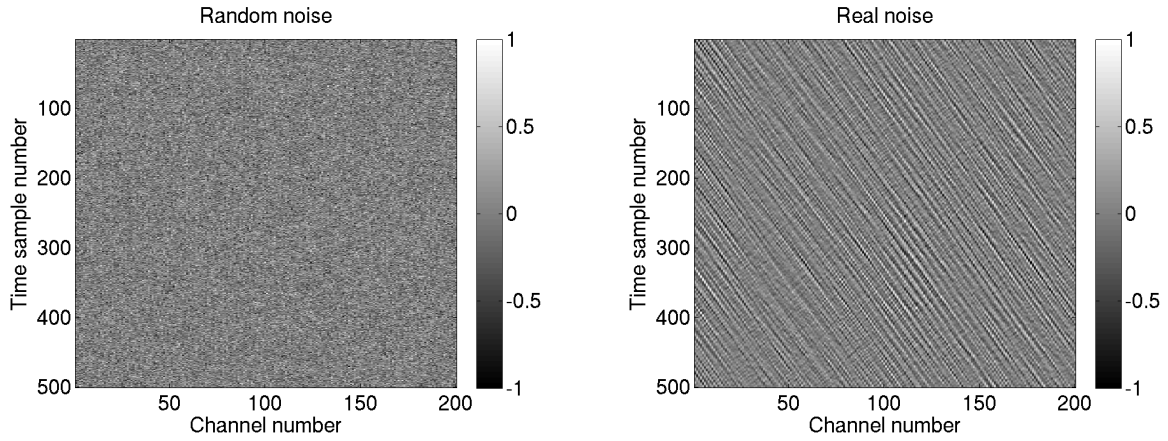


Figure 20: Noise samples. (left) Noise created from Gaussian random numbers. (right) Real noise dominated by tugging from the vessel.

dip.

The results from the tests are presented in Figures 21 - 24. Four methods were tested: correlation based method (Figure 21), structure tensor (Figure 22), linear plane wave destruction with the constraint (Figure 23) and nonlinear plane wave destruction (Figure 24). The left side of each of these figures presents tests with the dip value of 0.4 samples per trace and the right side presents tests with the dip value of 3.0 samples per trace. The upper figures show the estimated dip values $\hat{p}$ and their uncertainties Δp for both types of noise. The uncertainty is the standard deviation of the dip estimate. Its value Δp means that more than 70 $\%$ of the dip estimates fall in the interval $\hat{p} \pm \Delta p$. The middle figures show the standard deviations for different parameters when real noise was added while the lower figures show the standard deviations when random noise was added.

General observations

All the methods assume only one constant dip within the window of investigation and estimate the dip with the largest amplitude if more than one dip is present in the window. When SNR is low, the method starts catching the dip of the noise. The Gaussian random noise has mean value $\mu = 0$, therefore, when SNR is low, the dip estimate shifts to zero from its actual value (see for example, Figure 21, first plot). The real noise is dominated by fairly coherent tugging noise from the vessel. Its dip is larger than 3 samples per trace. Therefore, when SNR is low, the estimated dip is shifted towards the dip of the noise. Standard deviation values always increase as the SNR decreases. It is interesting that the level of SNR, not the type

of noise and not the parameters of a method play the important role. The methods have different sensitivity to noise, that is they fail at different SNR values.

When the SNR is high, all methods correctly estimate the dip value of 0.4. The dip value of 3.0 is correctly estimated by the correlation based method and the nonlinear plane wave destruction method while the structure tensor estimation and the linear plane wave destruction methods overestimate it. The latter two methods both share the intermediate step of gradient estimation which is possibly responsible for this overestimation.

Correlation based method

The correlation based method is a scanning method - it tries all dip values in the given dip scanning interval. The maximum scanned dip values as well as the number of interpolation points are set by the user. The smaller the range of expected dips the faster the evaluation, because fewer dip values need to be scanned. Normally, interpolation of data is necessary because the sampling in time is too coarse. The number of interpolation points limits the accuracy of the dip estimate. Here, it was set to 30, so the precision is 1/60 samples per trace. The method requires a temporal window which is large in comparison to other methods (a few tens of samples). These requirements make the method slow and of limited accuracy, thus, not suitable for large amounts of data. The results for different SNR and different temporal window sizes are shown in Figure 21. The method is still able to estimate the acctual dip value when SNR $= 5$. A larger temporal window size slightly reduces the standard deviation value. Standard deviations are smaller for real noise than for random noise, especially for the test dataset with dip of 3 samples per trace. This might be because the real noise contains dip close to 3 samples per trace.

Structure tensor estimation

The structure tensor estimation method is based on gradient calculations. It is important what kind of filter is used to estimate the gradient (see Section 6.2 for discussion). The method is also relatively sensitive to both real noise and Gaussian random noise. The window size reduces the standard deviation but it does not help to prevent the failure to estimate dip. The method is said to have failed when instability occurs and standard deviations become unreasonably large ($n\times 10^2$). In

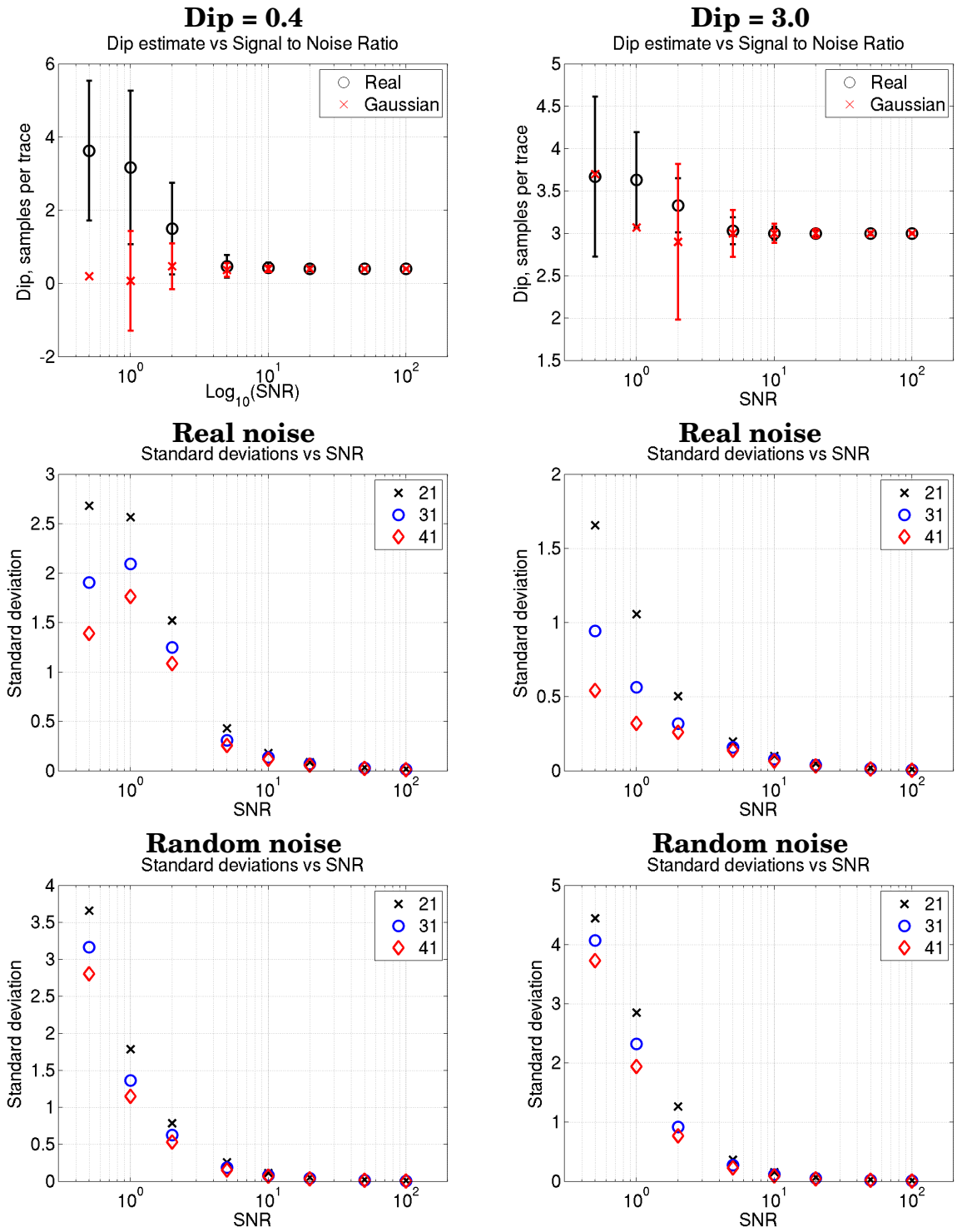


Figure 21: Tests of correlation based method. The temporal window sizes were respectively 21, 31 and 41. Standard deviation values larger then 2 are not shown in the upper figures. Logarithmic scale was used for SNR.

all of these tests, we define the standard deviation as being large when it is larger than 2 samples per trace. It can be seen from the tests that the structure tensor

method usually fails when SNR is equal to 10 or even 20 (Figure 22). Better results were achieved with bigger window sizes: 11 and 15 samples. These windows are big in comparison to the ones used in the plane wave destruction method and, therefore, unable to adapt to sudden dip changes in the data. So, though the computation time is short, successful results require a good dip filter and relatively high SNR.

Linear plane wave destruction method

The linear plane wave destruction method with the constraint (Section 3.4) is relatively fast and relatively insensitive to noise. It is important what type of dip filter is used for gradient calculation (see Section 6.2 for discussion). The chosen operator size is only 2×2 but the sliding window size is usully bigger to get smoother and more robust estimates. Here, the used window sizes were 4×4, 7×7 and 11×11, respectively (Figure 23). The window size can be small in comparison to the window sizes employed in the correlation and structure tensor methods, so the method is able to adapt to local dip changes in the data. It is interesting, that despite reducing the standard deviation, larger windows do not prevent the method from failure. It starts failing at SNR $= 5$ and, thus, has almost the same sensitivity to noise as the correlation based method. The method is said to have failed when the estimated dip value is incorrect and standard deviation is large. The strength of this method is that when it fails, the standard deviations are smaller than when the other methods fail: it is always up to 2 samples per trace and smaller for the random noise. So even the first failed values might be used as rough estimates.

Nonlinear plane wave destruction method

The nonlinear plane wave destruction method is relatively fast and insensitive to noise (Figure 24). The method employs the smallest dip operator (3×2 or 5×2, see Section 3.4) and, thus, has the highest ability to adapt to the dip changes in the data. On the other hand, some smoothness has to be applied to obtain physically acceptable results, because dips usually do not change abruptly. This is similar to applying a larger window. Smoothing also helps to get more reliable estimates – the standard deviation of dip becomes smaller as the smoothness parameter is increased. Another apparent drawback of the method is the need of an initial dip value. Smoothing is helpful here because even small values of the smoothing parameter eliminate the initial dip value's influence on the final result. Therefore,

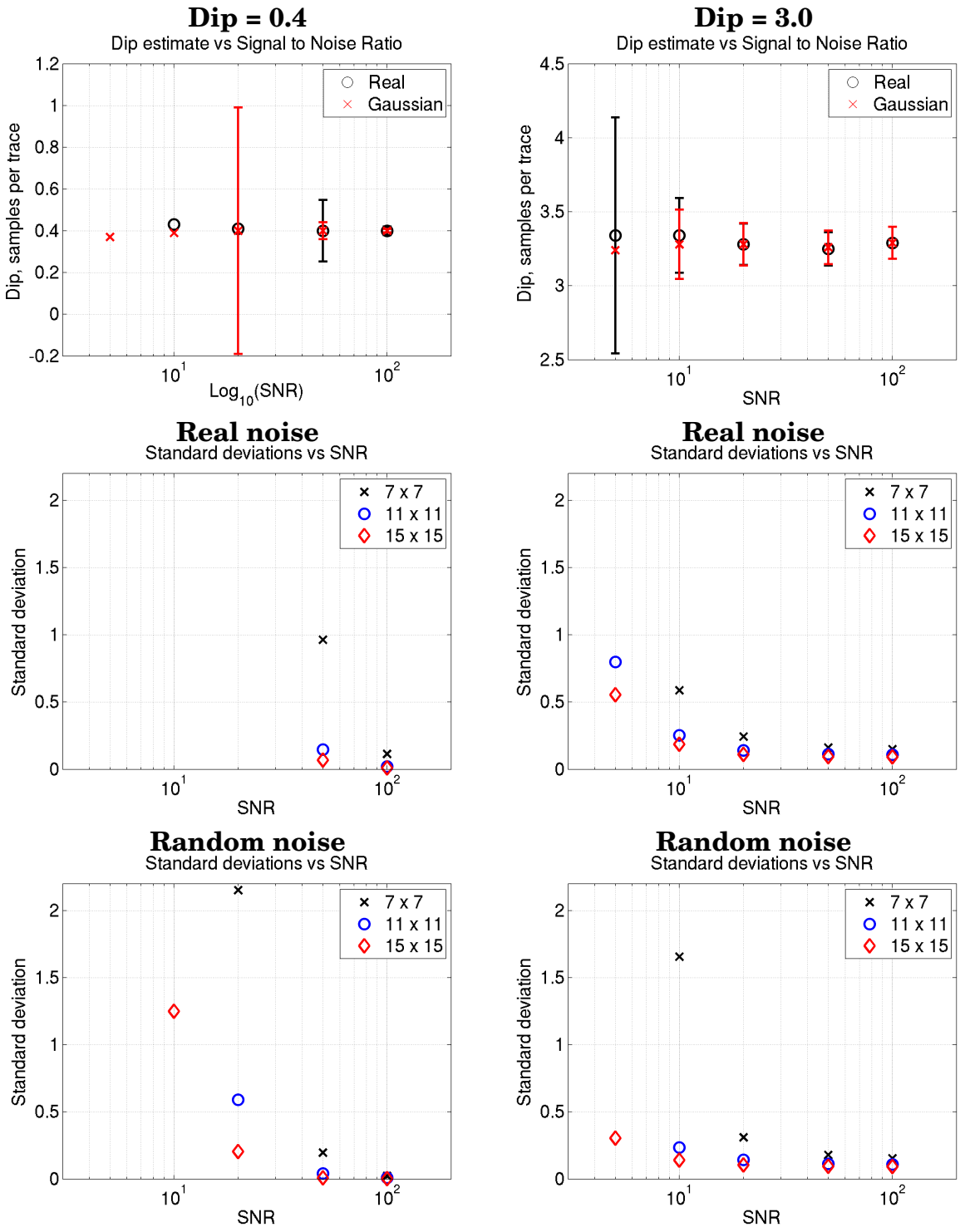


Figure 22: Tests of structure tensor estimation method. The sliding window sizes were respectively 7×7, 11×11 and 15×15. Some dip estimates and standard deviation values larger than 2 are not given because the tests failed (see the text for more detailed explanation). Logarithmic scale was used for SNR.

the initial dip value does not need to be accurate and, thus, can be obtained after the initial inspection of data. If smoothing is not desired we need a good initial dip

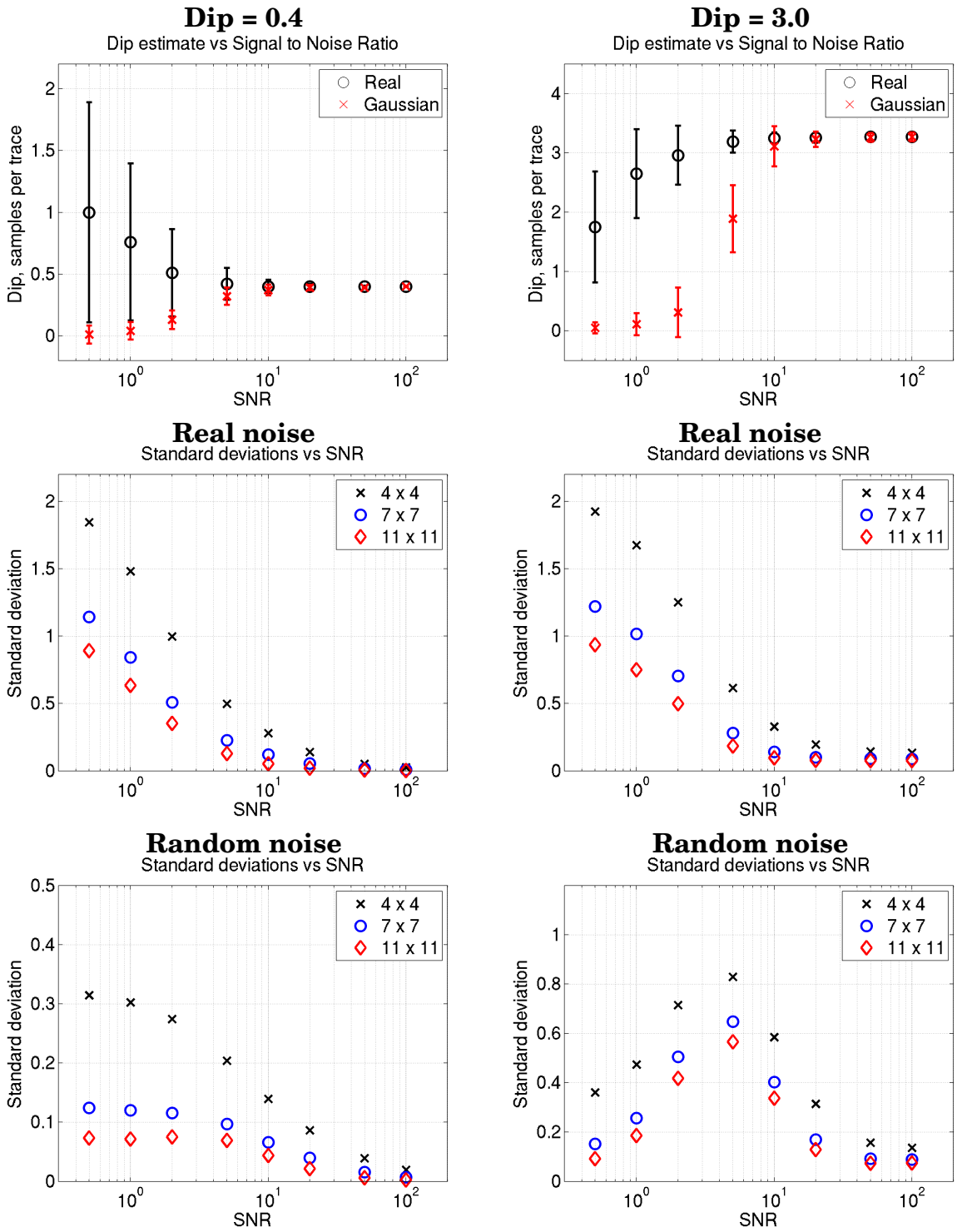


Figure 23: Tests of the linear plane wave destruction method with constraint. The sliding window sizes were respectively 4×4, 7×7 and 11×11. Logarithmic scale was used for SNR.

value, and dips can be estimated in an iterative manner. First, a fast dip estimation (one iteration) is run. Then, its output is used as initial dip values for further

iterations. The method estimates the actual dip values correctly with SNR $= 5$ and almost correctly even with SNR $= 2$. Thus, this method is the most insensitive to noise. The strength of this method is that when it fails, the standart deviations are smaller than when other methods fail: they are always up to 2 samples per trace.

Choise of optimal parameters

Since the parameters, such as window size, do not seem to change the SNR threshold of a method's failure but only change the standard deviation (reliability), the optimal values were chosen as a trade-off between reliability, computational time and ability to adapt to local changes of dip. For the correlation based method, the time window size can be set to 31 samples. For the structure tensor method, the window size can be set to 11×11 samples, and the finite difference stencil proposed by Claerbout can be employed. Weights can be chosen as described in Section 3.2. In case of the linear plane wave destruction method, the window size can be 7 samples and the same finite difference stencil can be used. In case of the nonlinear plane wave destruction method, the stencil size is chosen with respect to the maximum expected dip value and smoothing parameter can be set to 5. These conclusions are valid for data with temporal sampling interval of 2 ms. Other parameters might work better for other sampling intervals.

6.1.2 Tests on Real SI Example

This section presents tests on real data containing SI noise. The first test dataset was taken from the part of a shot gather where only noise is recorded, that is, before the direct and refracted arrivals (Figure 25, left). The second test dataset was taken from the late recording times of a shot gather (Figure 25, right), where the reflections are relatively weak. SI noise is clearly visible in both subsets. Moreover, the SI noise present in each of these subsets has nearly constant dip. Therefore, the same strategies as used in Sections 6.1.1 and 4 can also be applied for these tests.

Figures 26 - 28 show the results of dip estimation employing the linear plane wave destruction method for the first subset (Figure 25, left). Figure 26 shows the characteristics of the dip estimate (see Section 4): from left to right, the estimated dip

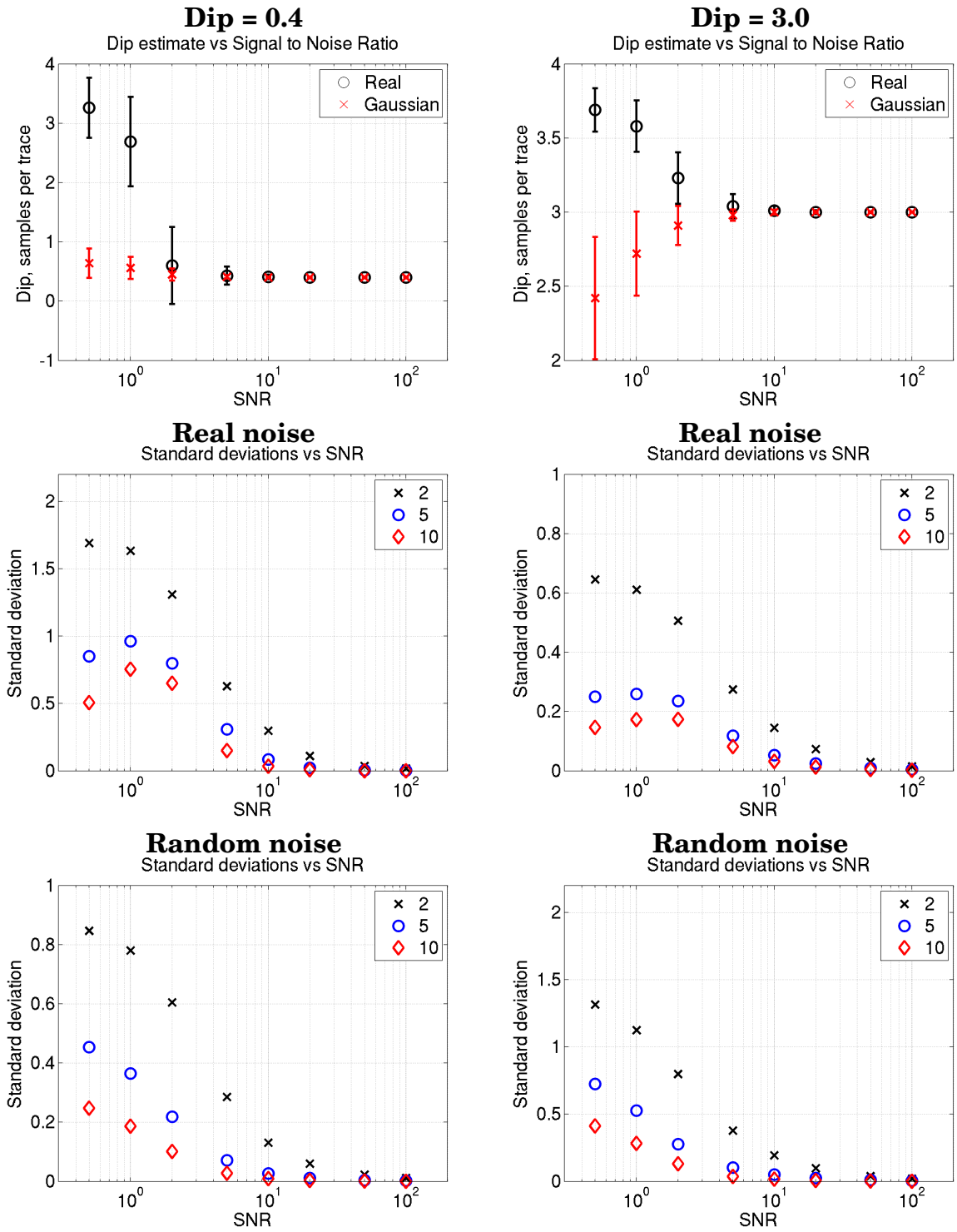


Figure 24: Tests of the nonlinear plane wave destruction method. The smoothness parameter values were respectively 2, 5 and 10. Logarithmic scale was used for SNR.

field, the histogram of this dip field, the local standard deviation and its histogram.

Dip estimators also produce an attribute - coherency measure, such as correlation

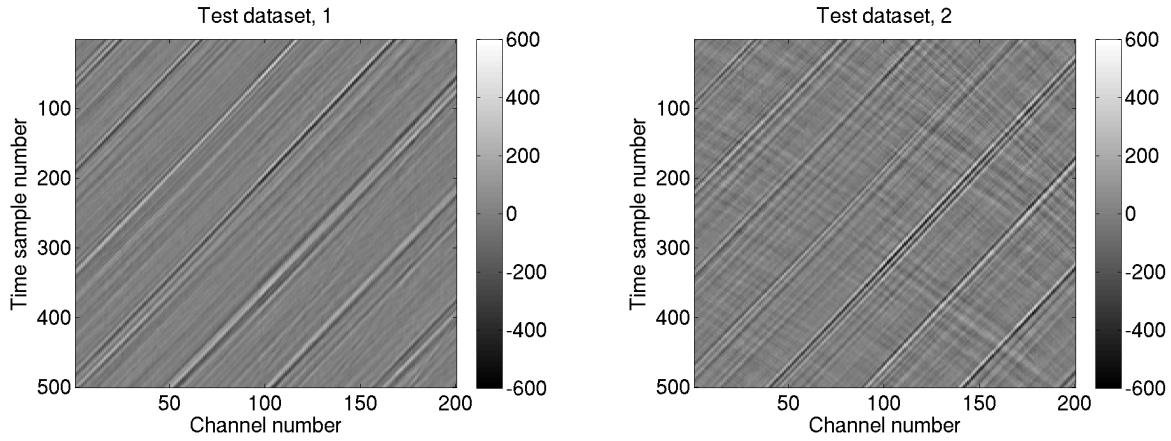


Figure 25: Test datasets. (left) A subset of a shot gather before direct and refracted arrivals. (right) A subset of a shot gather taken at late recording times.

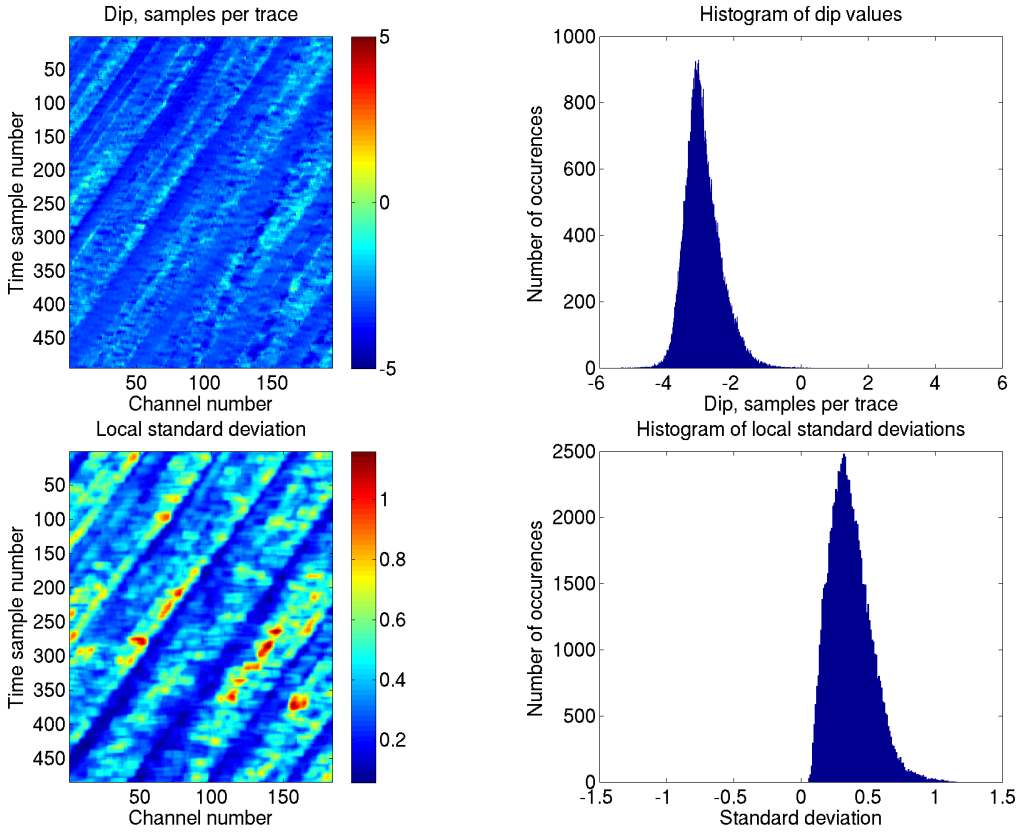


Figure 26: (from left to right) The estimated dip field, the histogram of this dip field, the local standard deviation and its histogram for the dataset in Figure 25, left.

coefficient, coherency or anisotropy. Together with the local standard deviation, they are used to select the reliable dip estimates. Figure 27 shows such an attribute - coherency together with the RMS values in μbar (the RMS window size was the same as the sliding window size in the dip estimator). The coherency attribute shows how good the dip was estimated. In the case of the plane wave

destruction method it is equal to one everywhere and, thus, is not helpul to identify the reliable dip estimates. RMS values may also be considered, because if the amplitude is weak the SI or any other signals are simply not present.

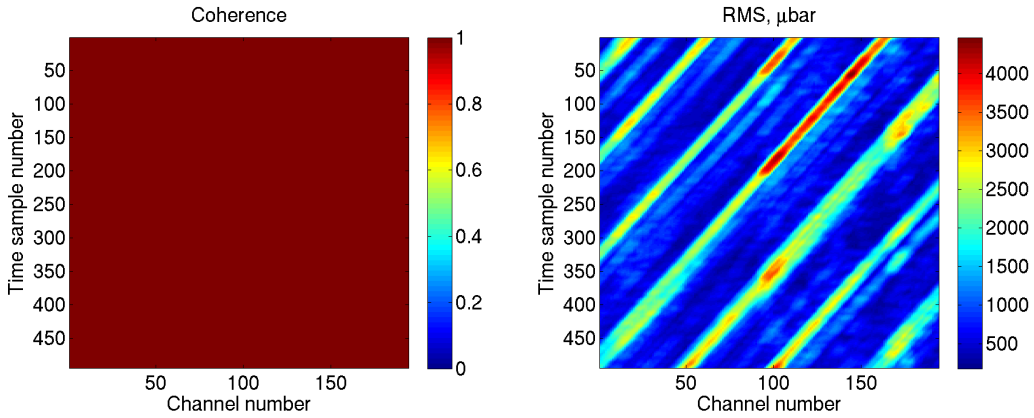


Figure 27: (left) Coherency attribute. (right) RMS values.

Figure 28 shows the estimated dip field and the selected reliable dip values. These values were selected by choosing the RMS amplitude to be larger than 1500 μbar and the local standard deviation to be less than 0.3. In cases where the coherency measure also varies (the correlation based method and the structure tensor estimation method), the reliable dip estimates have a coherency larger than 0.9. It was found, that these values are valid for all methods and the percentage of good dip estimates is about 40% for all the methods for the first dataset and about 30% for the second dataset (see Tables 2 - 5). There are fewer reliable dip estimates in the second dataset because here the weak reflection signals manifest as noise.

The considered dip estimation methods show some differences. For example, the structure tensor estimation technique produces two coherency measures: one called coherency and the other anisotropy (Section 3.2). However, it was noticed that the coherency attribute is more sensitive and, thus, more helpful for dip evaluation. The linear plane wave destruction method also produces the residual representing the original dataset after the estimated dip has been removed. As it was mentioned in this section, coherency measure is not helpful in case of plane wave destruction methods. As we see from Figure 29, the SI noise has been succesfully attenuated and the residual amplitudes are low. However, the residual can only be used for visual inspection because it represents some kind of directional derivative.

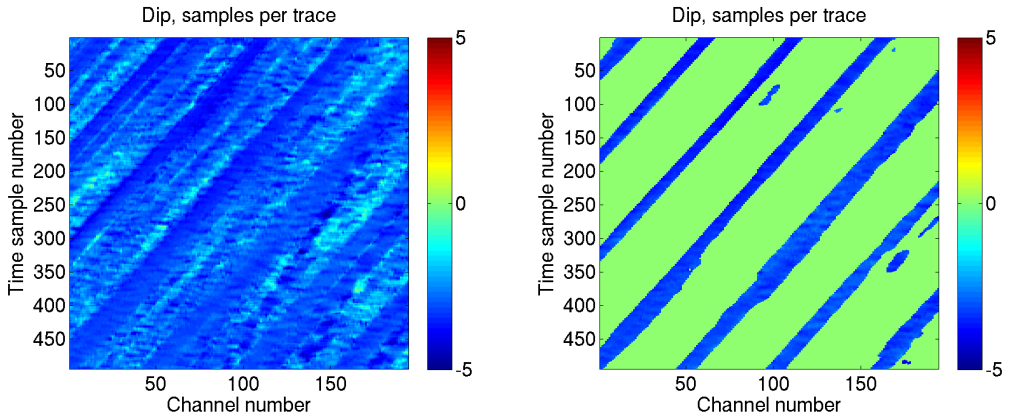


Figure 28: (left) The estimated dip field. (right) The reliable dip estimates. In this presentation, zero values indicate unreliable dip estimates.

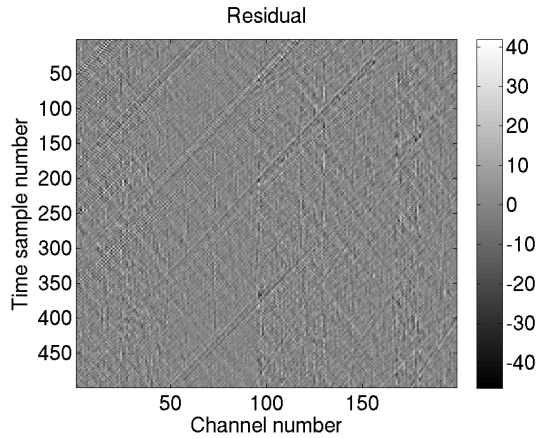


Figure 29: Residual after removal of the estimated dip (the SI noise dip) by the linear plane wave destruction method.

The results obtained from testing with other methods as well as the results for the second dataset are presented only in tables. Table 2 presents the results for the first dataset and Table 3 for the second dataset. The optimal parameters for each method were chosen based on the previous tests performed on synthetic data (Section 6.1.1).

It seems that the methods identify slightly different dip values. The structure tensor method fails even for the first dataset (standard deviation is large), which is ”clean” from reflections. The correlation method gives similar dip estimate as the nonlinear plane wave destruction method. Since it is known that methods based on finite difference stencils (e.g., linear plane wave destruction and structure tensor estimation methods) might incorrectly estimate dips larger than one, the same test datasets can be transposed. Then, dips are clearly less then one. The results

Table 2: Dip evaluation results for the first dataset (Figure 25, left). STD stands for 'standard deviation', PWD for 'plane wave destruction', and PoGE for 'part of good estimtes'. The values in brackets show the estimates when only the reliable dip values were considered.

Method	Estimated dip	STD	Spatial STD	PoGE
Correlation	-2.80 (-2.82)	0.50 (0.15)	0.16	0.40
Structure tensor	-3.34 (-3.53)	— (0.51)	0.45	0.42
Linear PWD	-3.05 (-3.23)	0.52 (0.27)	0.32	0.37
Nonlinear PWD	-2.81	0.13	0.03	—

Table 3: Dip evaluation results for the second dataset (Figure 25, right). STD stands for 'standard deviation', PWD for 'plane wave destruction', and PoGE for 'part of good estimtes'. The values in brackets show the estimates when only the reliable dip values were considered.

Method	Estimated dip	STD	Spatial STD	PoGE
Correlation	-2.83 (-2.77)	2.57 (0.55)	0.22	0.32
Structure tensor	-3.83 (-4.09)	— (26.77)	0.46	0.26
Linear PWD	-3.05 (-2.89)	1.45 (0.91)	0.61	0.27
Nonlinear PWD	-2.81	0.84	0.08	—

of dip estimation of the transposed datasets are presented in Table 4 for the first dataset and Table 5 for the second dataset.

Standard deviations are larger for the second dataset because it also contains weak reflections. In this case, they act like "noise" because these events are much weaker than the SI noise. The estimated dip values depend on the method and can be divided into two groups: the estimates obtained from the correlation based and the nonlinear plane wave destruction methods, and the estimates obtained from

Table 4: Dip evaluation results for the first dataset transposed (Figure 25, left). STD stands for 'standard deviation', PWD for 'plane wave destruction', and PoGE for 'part of good estimtes'. The values in brackets show the estimates when only the reliable dip values were considered.

Method	Estimated dip	STD	Spatial STD	PoGE
Correlation	-0.33 (-0.34)	0.03 (0.02)	0.02	0.66
Structure tensor	-0.28 (-0.29)	0.06 (0.03)	0.04	0.27
Linear PWD	-0.28 (-0.29)	0.05 (0.02)	0.03	0.23
Nonlinear PWD	-0.35	0.33	0.03	—

Table 5: Dip evaluation results for the second dataset transposed (Figure 25, right). STD stands for 'standard deviation', PWD for 'plane wave destruction', and PoGE for 'part of good estimtes'. The values in brackets show the estimates when only the reliable dip values were considered.

Method	Estimated dip	STD	Spatial STD	PoGE
Correlation	-0.33 (-0.28)	0.10 (0.08)	0.04	0.76
Structure tensor	-0.26 (-0.25)	0.15 (0.05)	0.09	0.26
Linear PWD	-0.26 (-0.24)	0.13 (0.07)	0.06	0.27
Nonlinear PWD	-0.31	1.00	0.04	—

the structure tensor and the linear plane wave destruction methods.

The dip values of the non-transposed and the transposed datasets should be related by: $p \times p^{(T)} = 1$, where $p^{(T)}$ is the dip of the transposed dataset. Only the reliable dip estimates are considered. The dips of the original and transposed datasets multiply into a number very close to one for the first dataset and usually less than one for the second dataset. Unfortunately, this consideration does not help choose the best method.

Regardless the dataset, the estimated dip values are similar for the correlation based method and the nonlinear plane wave destruction method. They are also similar for the structure tensor estimation method and the linear plane wave destruction method. The former two methods are completely different in theory, while the latter two methods share the same method of gradient estimation (the same stencil) as an intermediate step. This observation may suggest that the dip estimates obtained from the latter two methods are biased because of the gradient estimation. Then, it can be concluded that the former two methods estimate dip more correctly. Moreover, this conclusion can be supported by the results of the tests in Section 6.1.1, where both correlation based and nonlinear plane wave destruction methods estimated the dip correctly for both tested dip values while both structure tensor estimation and linear plane wave destruction methods overestimated the dip value of 3.0. However, it is still not clear why dip estimates are different for the transposed datasets.

Table 6: Summary of the properties of tested methods.

Method	Accuracy	Insensitivity	Runtime
Correlation	Poor	Good	Very slow
Structure tensor	Average	Bad	Fairly fast
Linear PWD	Good	Good	Fast
Nonlinear PWD	Very good	Good	Fast

Table 7: Summary of the parameters to consider for tested methods.

Method	Parameters to consider
Correlation	Window size, number of interpolation points
Structure tensor	Window size, weights, stencil
Linear PWD	Window size, stencil
Nonlinear PWD	Smoothing, filter order (stencil)

6.1.3 Choise of the Best Dip Estimation Method

Tables 6 and 7 summarize the main observations made about the methods. By accuracy we mean the ability of the method to adapt to local changes of dip. The insensitivity to noise is desired, that is, the ability of the method to reliably and correctly estimate dip in the presence of different levels and types of noise. Only a qualitative estimate of runtime is given, because the accurate CPU time depends on, for example, the machine used, the implementation, the programming language and so on. Moreover, the quantitative comparison is not meaningful because the Madagascar package implementation of the nonlinear plane wave destruction was used, while the rest of the methods were implemented in MatLab as part of this thesis.

The nonlinear plane wave destruction method is chosen as the best one because it correctly estimates dip, the estimated dip is reliable (standard deviation values are relatively low even when the method has already failed (Section 6.1.1)), it adapts to the local changes in dip because it uses a small stencil, the method is relatively insensitive to noise - it does not fail even with SNR values as low as 5 or even 2 and, finally, it is fast and suitable to process large amounts of data.

The correlation method is rejected because it is slow and unsuitable to process large amounts of data. The structure tensor estimation method is problematic because of poor performance in the presence of noise. The linear plane wave destruc-

tion technique suffers from the dispersive properties of finite difference stencils though it can be a good solution if steep dips are not expected and / or the data does not contain high frequencies.

6.2 Choise of the Optimal Dip Filter

Ideal dip filter

An ideal dip filter removes a dip in the T−X domain or a certain velocity $v_x = f/k_x$ in the F−K domain. Thus, to understand how different T−X domain dip filters work, we can map them to F−K domain. This section discusses the properties of a few dip filters mentioned in this work.

Finite difference stencils

Finite difference (FD) stencils are symmetric, that is, they can estimate zero dips as well as infinite dips. However, the latter is not necessary in seismic data anlysis because dips are finite. As mentioned in Section 2.4, they are limited up to 4.16 samples per trace for 2 ms shot domain data used in this work.

Figures 30 and 31 show F−K spectra of some finite difference stencils. These spectra were calculated by applying the 2-D Fourier transform for the dip filters in T−X domain. Frequencies and wavenumbers are normalized. For all these spectra, blue color represents damped frequencies and red color - preserved frequencies. The F−K spectra of these stencils are presented: first order forward difference stencil without any smoothing and the same with smoothing (Claerbout's stencil, size 2 $\times$ 2 samples) in Figure 30, and second order central finite difference stencil with smoothing (Sobel operator, size 3 $\times$ 3 samples) and fourth order central finite difference stencil with smoothing (size 5 $\times$ 5 samples) in Figure 31.

The stencils dampen a single phase velocity at low frequencies, that is, they behave as good dip filters at low frequencies. Claerbout's stencil seems to be most linear in character and the one which attempts to remove only one velocity without damping any frequencies. It is also small in in size, only 2 $\times$ 2 samples. However, the larger dip, the smaller the frequency interval at which the dip filter's behaviour is linear. The forward FD stencil without smoothing boosts high frequencies while central FD stencils damp them. The larger the central FD stencil, the more of

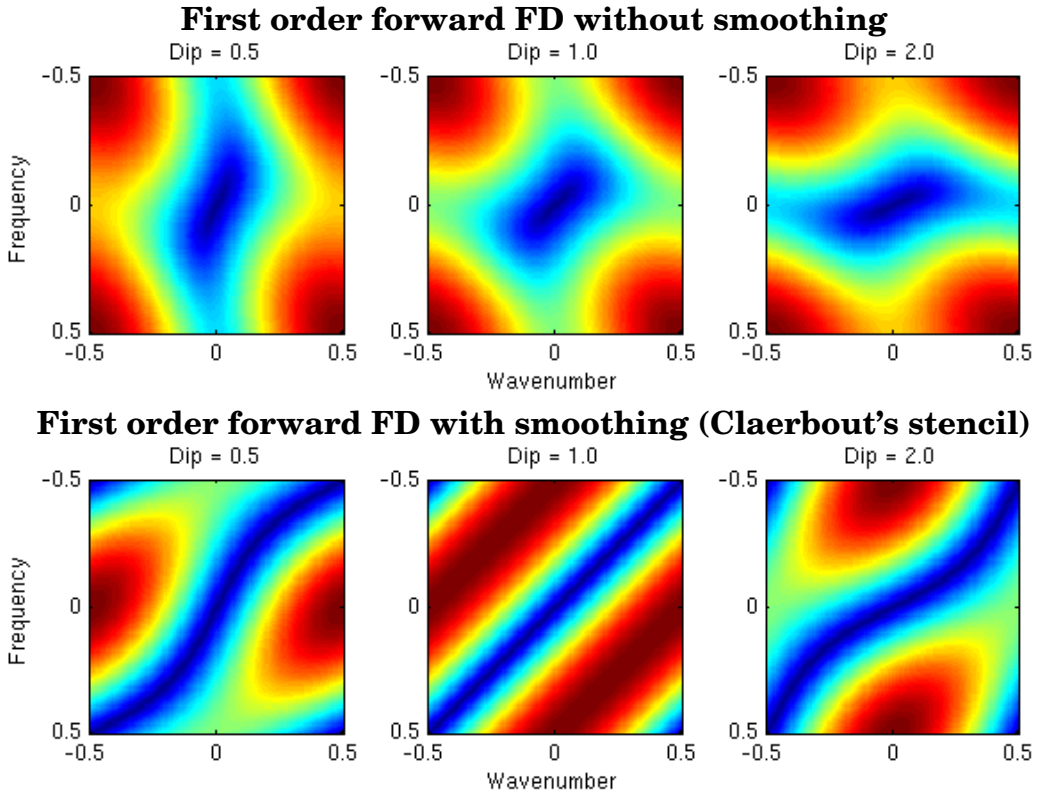


Figure 30: Normalized F−K spectra of various finite difference stencils for dip values 0.5, 1.0 and 2.0 samples per trace. (upper) First order forward finite difference stencil without smoothing. (lower) First order forward finite difference stencil with smoothing (Claerbout's stencil).

high frequencies it damps and the desired behaviour is observed only at very low frequencies. To summarize, the FD stencils are dispersive at higher frequencies and the larger they are, the lower frequencies get damped.

To make the finite difference stencil based gradient methods work, the aliased frequencies can be filtered. However, in the case of steep dips only the lowermost frequencies would be used. Although it was not tested, high-cut filtering might change the data to such an extent that the actual dip values cannot be recovered.

Gaussian derivative filters

The derivative of a Gaussian function can be used as a gradient operator. Such an operator is symmetric as finite difference stencils, thus can estimate any dip value. However, in order to obtain reliable estimates, the Gaussian derivative stencils need to be quite large, 10+ samples. Large stencils might smear the dip estimate. The parameters which can be varied are the standard deviations in both gradient

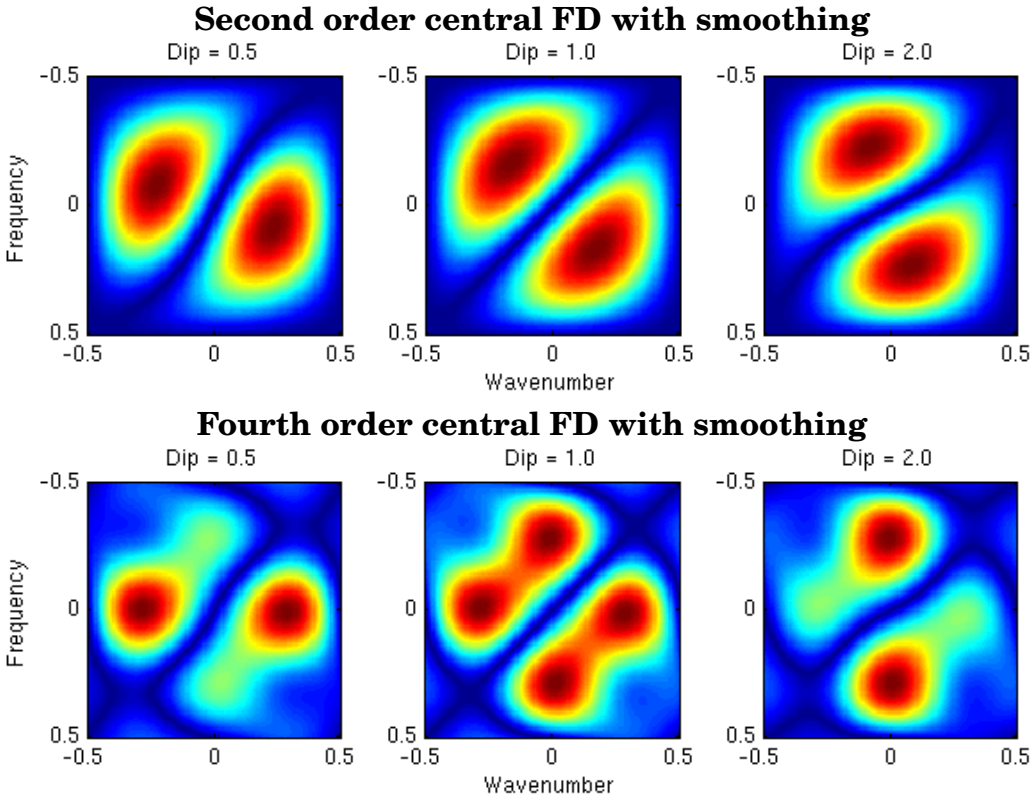


Figure 31: Normalized F−K spectra of various finite difference stencils for dip values 0.5, 1.0 and 2.0 samples per trace. (lower) Second order central finite difference stencil with smoothing (Sobel operator). (lower) Fourth order central finite differencing stencil with smoothing.

and smoothing directions. Figure 32, top, shows the F−K spectrum of the Gaussian derivative filter with standard deviation values of 3 in the gradient direction and 1 in the smoothing direction. The spectrum shows that such a filter damps a lot of frequencies and wavenumbers. It has the desired dip filter behaviour for a very small interval of low frequencies and low wavenumbers. Thus, it is not clear what it does to data and it is not suitable for our purposes. Figure 32, bottom, shows the F−K spectra of the Gaussian derivative filter with standard deviation values of 1 in both gradient and smoothing directions. Now, the spectrum clearly defines a velocity (dip) but still damps a lot of higher frequencies and wavenumbers.

Fomel's filters

Prediction error filter based stencils are asymmetric. However, as it was mentioned, symmetric filters are not necessary, if very steep dips are not expected. The F−K spectra of 3- and 5- point stencils without the constraint (Section 3.4)

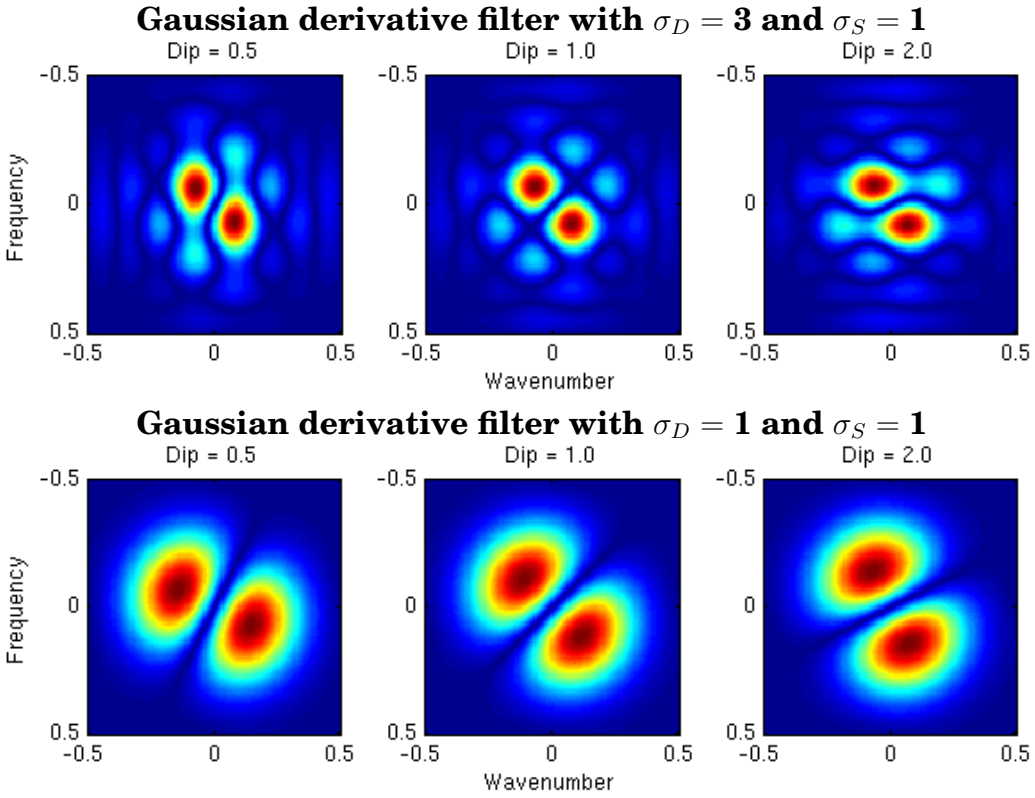


Figure 32: Normalized F−K spectra of the Gaussian derivative filters for dip values 0.5, 1.0 and 2.0 samples per trace. (upper) The standard deviation value of 3 in the direction of gradient and 1 in the direction of smoothing. (bottom) The standard deviation values of 1 for both gradient and smoothing directions.

are given in Figure 33. Both filters are aliased for dip value of 2.0, and the larger filter damps more high frequencies.

The F−K spectra of the constrained filters are given in Figure 34. These filters are not aliased, and the desired behaviour of a dip filter is observed with the dip value of 2.0 as well. These filters are said to have the antialiasing property. The larger filter damps more high frequencies, as it is in the case of the unconstrained filter. Prediction error based stencils are small (3×2 and 5×2), thus able to adapt to dip changes. Larger stencils are capable of estimating or removing larger dips and can be constructed if large dips are expected 3.4. All these stencils include only two traces, thus, dips are always strictly planar while this condition is only approximate for other stencils which include more traces (have larger window size in the spatial direction).

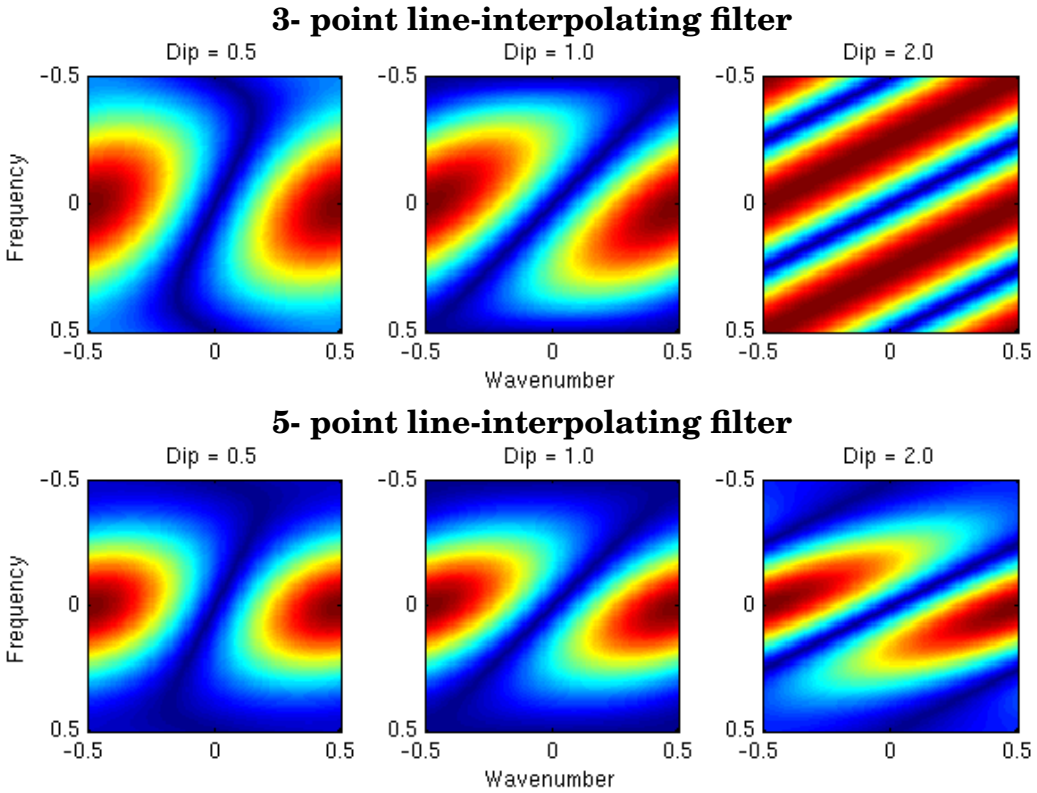


Figure 33: Normalized F−K spectra of the unconstrained 3- (top) and 5- (bottom) point filters for dip values 0.5, 1.0 and 2.0 samples per trace.

6.3 Application Example: Dip Estimation in Seismic Line

The dips of SI noise were estimated for one seismic line. The nonlinear plane wave destruction method was used because it was chosen as the best. Evaluation of the dip estimates is carried out in the way presented in Section 4.

An example of such estimation, for one shot gather from the seismic line used in this work, is presented in Figure 35. The input shot gather is shown to the left and its dip field in the middle. In this example, the number of occurences of dip values has two peaks for each trace (Figure 35, right). The peak with the positive dip value is that of the signal and the peak in the negative dip interval represents the SI noise. This dip is almost the same for all traces. The number of occurences of the SI noise dip is small for the first traces of the shot gather because reflections dominate.

It can be seen from Figure 35 that the dip of the SI noise is independent of the trace number. Therefore, dip estimation (see Section 4) for the seismic line used

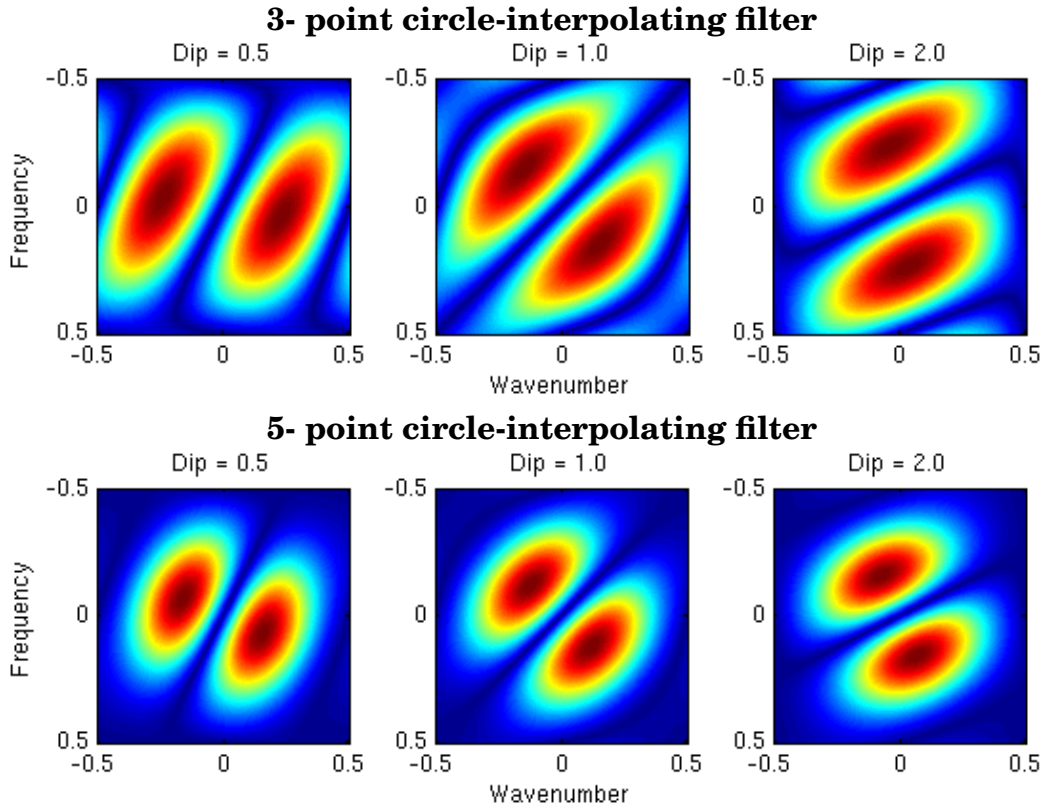


Figure 34: Normalized F−K spectra of the constrained 3- (top) and 5- (bottom) point filters for dip values 0.5, 1.0 and 2.0 samples per trace.

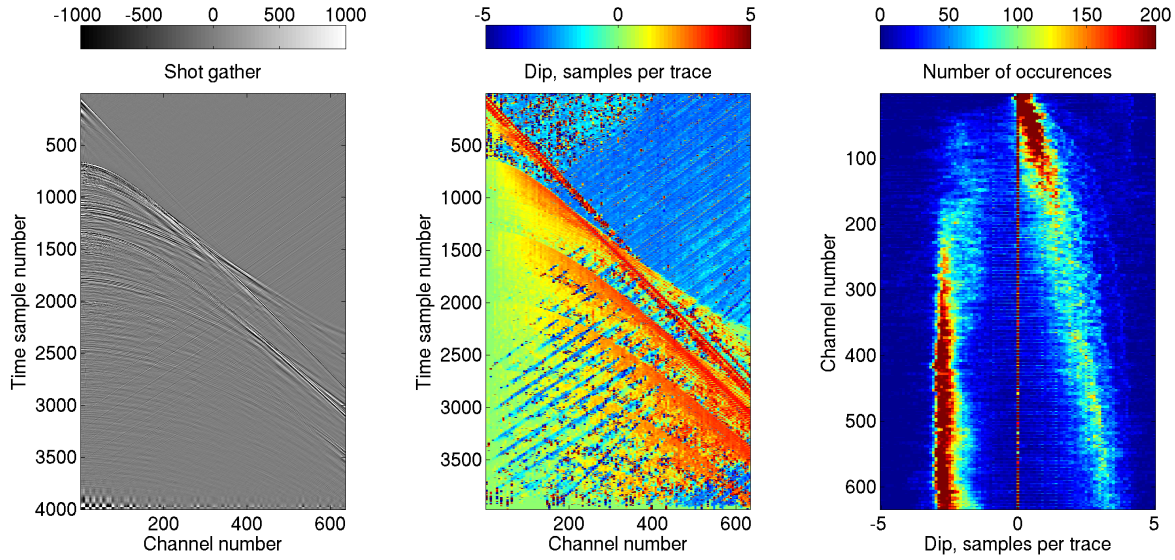


Figure 35: (from left to right) Input shot gather, the estimated dip and the number of occurences of dip values as a function of channel number and dip value.

in this work can be simplified. The function of number of occurences is reduced by one dimension and now depends on shot number and dip value. The most frequent

dip is the dip of the SI noise and only reliable values are accepted.

Since the arrival time of the SI noise train appears random in this seismic line (Figure 8), to accomplish the dip estimation it is sufficient to select a subset of data before the direct and refracted arrivals (an example is the first subset in Section 6.1.2, Figure 25, left). The dip of that subset is the SI dip assigned for the entire shot.

The final result of this estimation is presented in Figure 36. Dips were estimated for each shot and only reliable dip estimates were accepted as the SI dip values. Each dip estimate has attributes which determine its reliability, namely, standard deviation and the most frequent local standard deviation (see Section 4). Coherency attribute could not be used because in the case of plane wave destruction methods it is all equal to one (see Section 6.1.2).

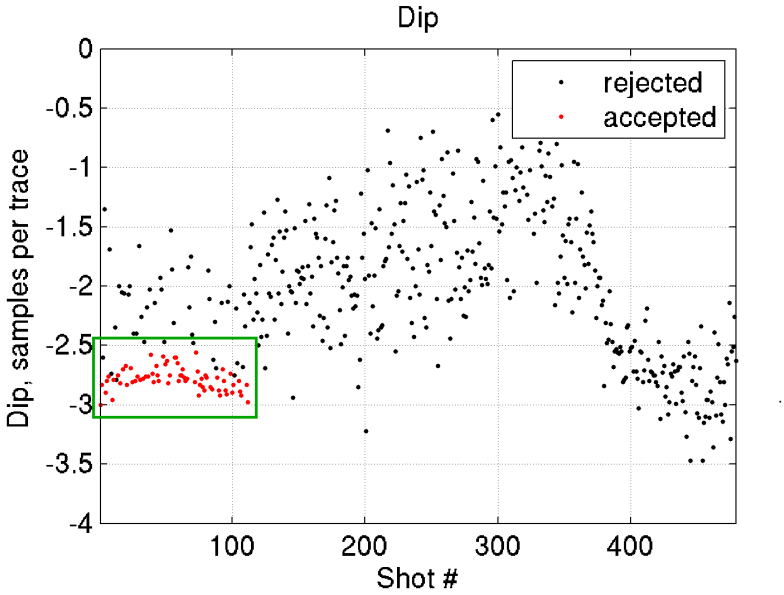


Figure 36: Results of SI dip estimation for the seismic line: the most frequent dip value as a function of shot number. The reliable dip values are shown by a green rectangle.

In this line, the SI noise is present in the first ∼110 shots (Figure 36). Since the arrival time of SI noise is random (see Figure 8), the noise could not be detected in all these ∼110 shots. Figure 37 shows the standard deviation as a function of shot number on the left and the crossplot of dip values against standard deviation on the right. Dip values with low standard deviation (here, ∼0.5) represent the

reliable dip estimates. Local standard deviation is not presented with a figure because it was very similar to standard deviation. Since the coherency could not be used, we present the RMS amplitude of each window as a function of shot number (Figure 38, left) as well as their crossplot (Figure 38, right). Here, high RMS amplitudes indicate the presence of SI noise and are expected only in the first ∼110 shots (left).

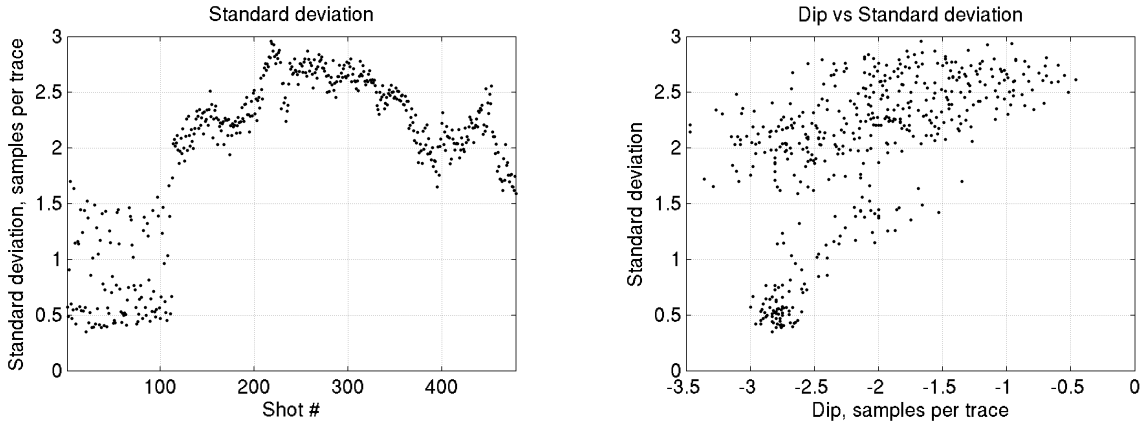


Figure 37: Results of SI dip estimation for the seismic line. (left) Standard deviation of dip as a function of shot number. (right) Crossplot of dip and standard deviation.

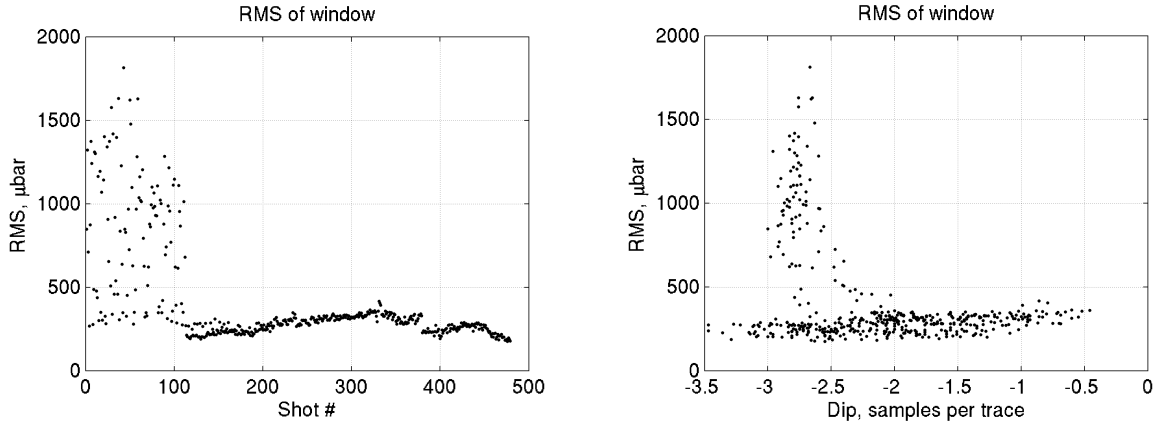


Figure 38: Results of SI dip estimation for the seismic line. (left) RMS of amplitude in the window as a function of shot number. (right) Crossplot of dip and RMS of amplitude in the window.

To summarize, in this seismic line the dip SI noise ranges from -3.0 to -2.6 samples per trace. It seems that there is a minor trend: dips approach -3.0 as the shot number increases (Figure 36). These values will be used in Section 6.4.1 as the input for SI attenuation procedure.

6.4 Application Example: SI Removal from Seismic Line

6.4.1 SI Attenuation by the Tau-p - Common-p Method

This subsection presents an example of how SI noise was removed from one seimic line by the $\tau - p$ - common$-p$ method. Figure 39 presents the main results of the SI removal for one shot. The input shot gather is separated into the signal part and the SI noise part. Raw input gathers were filtered to remove the low frequency components and swell noise. The SI noise in this line appears as trains of plane waves with negative dips (negative p values). Since the SI dip does not change from trace to trace, the noise source might have been as far as almost 100 km away. The signal part is obtained by adaptivelly subtracting the SI noise model from the input gathers. The SI noise was obtained by the $\tau - p$ common$-p$ method (Section 5.2). Though visually the signal part appears fine, the SI model (the noise part) contains not only SI noise, but some signal too. This might have happened because of the values of parameters used in the random noise attenuation routine being too tough.

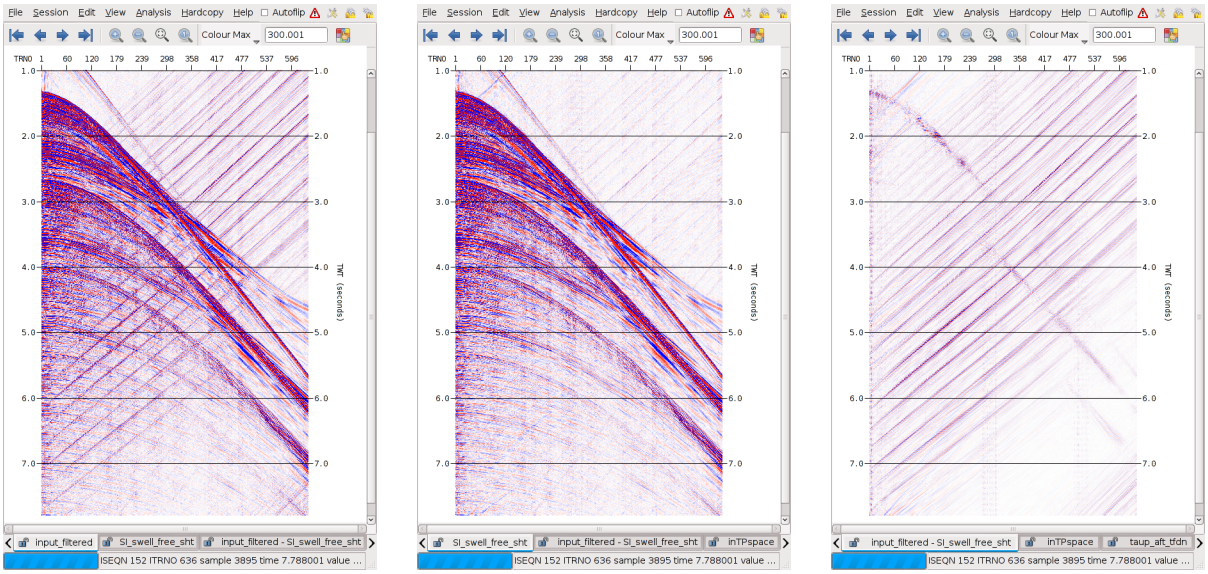


Figure 39: (from left to right) Filtered input shot gather, signal part, and difference between input and signal part (SI noise model).

The following figures present some intermediate steps of the $\tau - p$ - common$-p$ flow. Figure 40 shows the $\tau - p$ transforms of the shot gathers presented in Figure 39. Shot gathers were interpolated before $\tau - p$ transform because the transform is not fully signal preserving. After mapping into the $\tau - p$ domain, the SI noise train

(plane waves) is located at roughly the same moveout p value. The τ values are different because each plane wave in the noise train arrives at different zero-offset time.

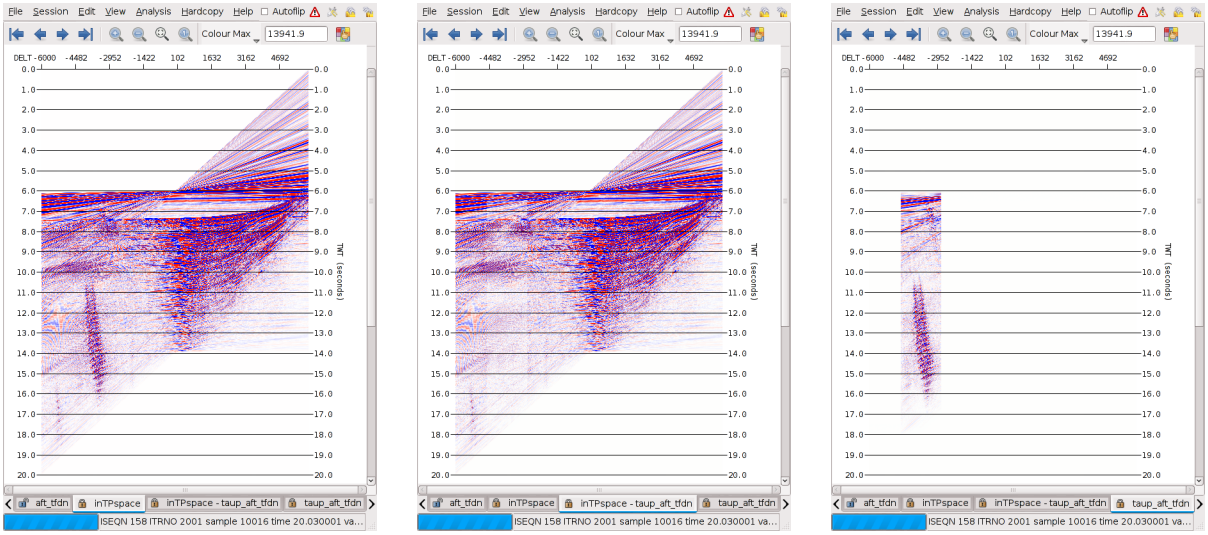


Figure 40: (left) $\tau - p$ transforms of the gathers shown in Figure 39.

The random noise attenuation routine was applied only in the p interval of the SI noise, as shown in Figure 40, right. This routine could also be applied on the entire p range, especially, if the SI dip range was not estimated before denoising. However, this takes much more computational time and might slightly damage the signal. Therefore, only the estimated SI dip range is used. Since the software requires moveout instead of dip in samples per trace, the unit conversion was made as described in Section 2.4. In order to be sure that all the SI noise would be attenuated, a larger dip SI range than estimated in Section 6.3 was chosen. The common$-p$ gather with SI before random noise attenuation is shown in Figure 41, left. The SI noise appears random in this domain. Figure 41, middle shows the same common$-p$ gather after random noise attenuation. Now, the common$-p$ gather appears cleaner. The difference of the original common$-p$ gather and the de-noised common$-p$ gather is shown in Figure 41, right. This difference is used to produce the SI model (Figure 39, right) as follows: it is sorted back to the $\tau - p$ domain and transformed back to the shot domain. The SI-noise attenuated data is obtained by subtracting the SI model from the original filtered shots (Figure 39, middle).

Figures 42, 43 and 44 present post-stack migrated seismic sections generated from

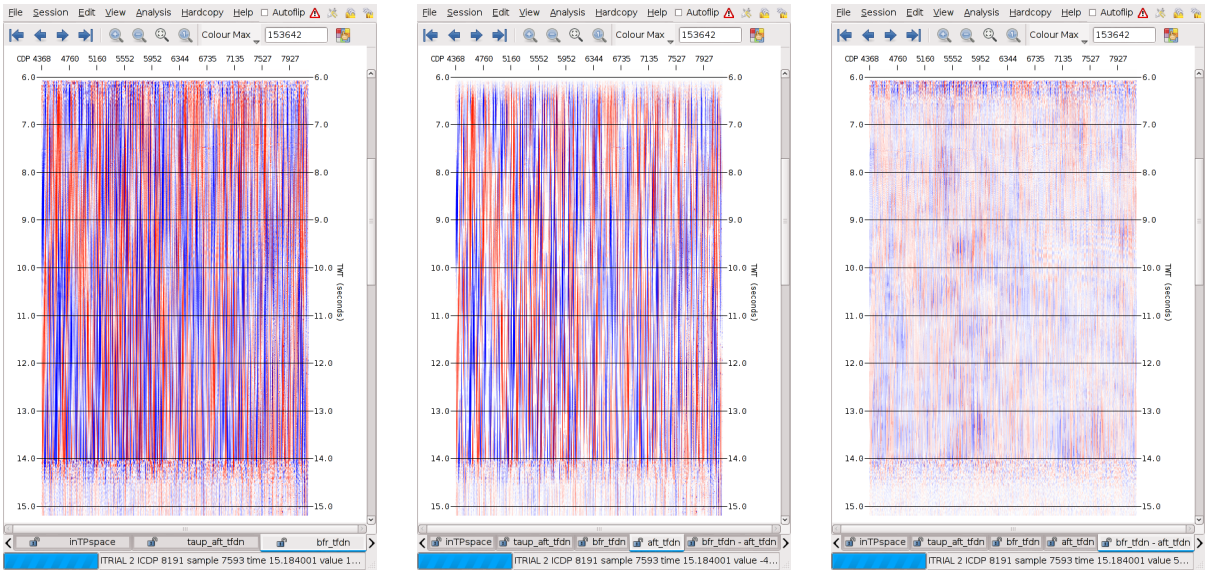


Figure 41: (left) A common$-p$ gather before random noise attenuation. (middle) The common$-p$ gather after random noise attenuation. (left) The difference common$-p$ gather.

SI contaminated data, SI-attenuated data and their difference, respectively. As can be seen from the difference image, the SI-contaminated and SI-attenuated images are almost identical. This is because the combined use of stacking and post-stack migration represent powerful denoising of data. By close inspection one can notice weak SI noise which appears as dense and steep linear events in the first $sim110$ CMP positions of the difference image. If pre-stack analysis like amplitude versus offset (AVO) is to be carried out or pre-stack migration is ought to be done, SI attenuation is still important.

6.4.2 Dip Based Signal and Noise Separation

This section presents the results from using dip-based signal and noise separation. As mentioned in Section 5.3, dips of signal and noise should be fairly different. In case of the seismic line used in this work, this condition is satisfied because the SI dip is negative and thus always different from that of the data.

The Madagascar package was used and two dip fields were simultaneously estimated using the function ”sftwodip2” and then separated using function ”sfplanesignoi”. Figure 45 shows the input data and the both dip fields estimated by ”twodip2”. Figure 46 presents the result of signal separation from noise. Visually,

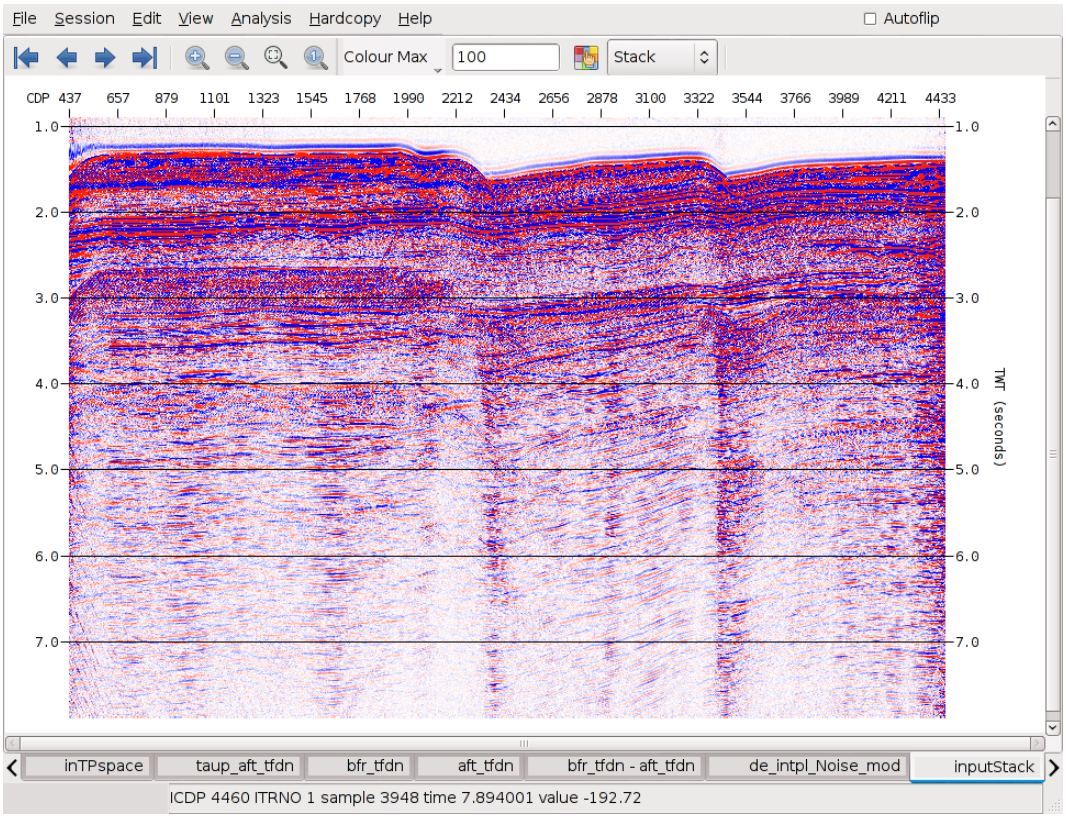


Figure 42: Migrated section based on original data (including SI).

the signal part looks well separated but the noise part contains too much signal. A better separation may be obtained by iterating the procedure a few times, however, the noise part still contains some signal (Figure 47). A possible problem is that the SI noise train is not really present in all parts of the shot gather. Therefore, there is no second dip to identify while the function ”twodip2” still tries to identify two dips and fails where effectively only the signal is present. This is clearly the case for the first traces and the water bottom reflection, where the signal is very strong.

In order for the method to be successful, the dip of the signal and the dip of the SI noise should be fairly different. Moreover, the SI contaminated areas of the shot gathers should be identified and the dip based signal and noise separation should be done only in those areas. This strategy can help to preserve the signal. In general, though it is an interesting possibility, these reasons make dip-based SI separation in shot gathers an unattractive solution.

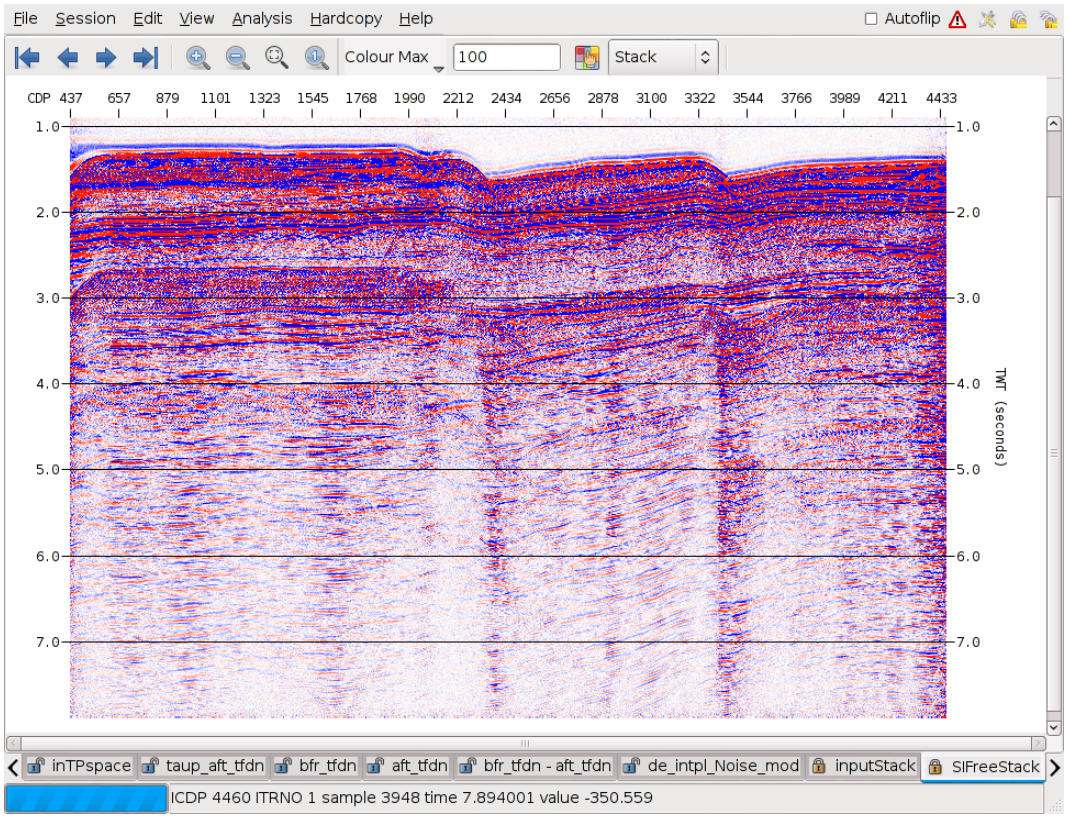


Figure 43: Migrated image based on SI-attenuated input data.

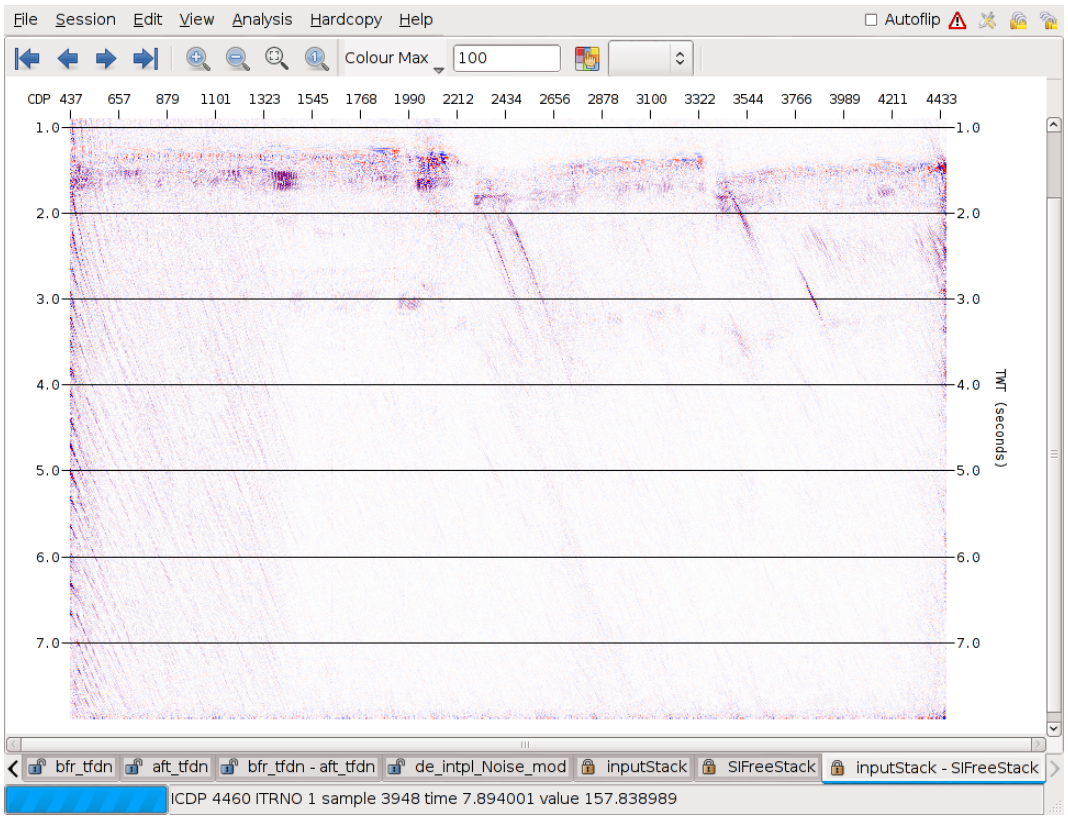


Figure 44: Difference image.

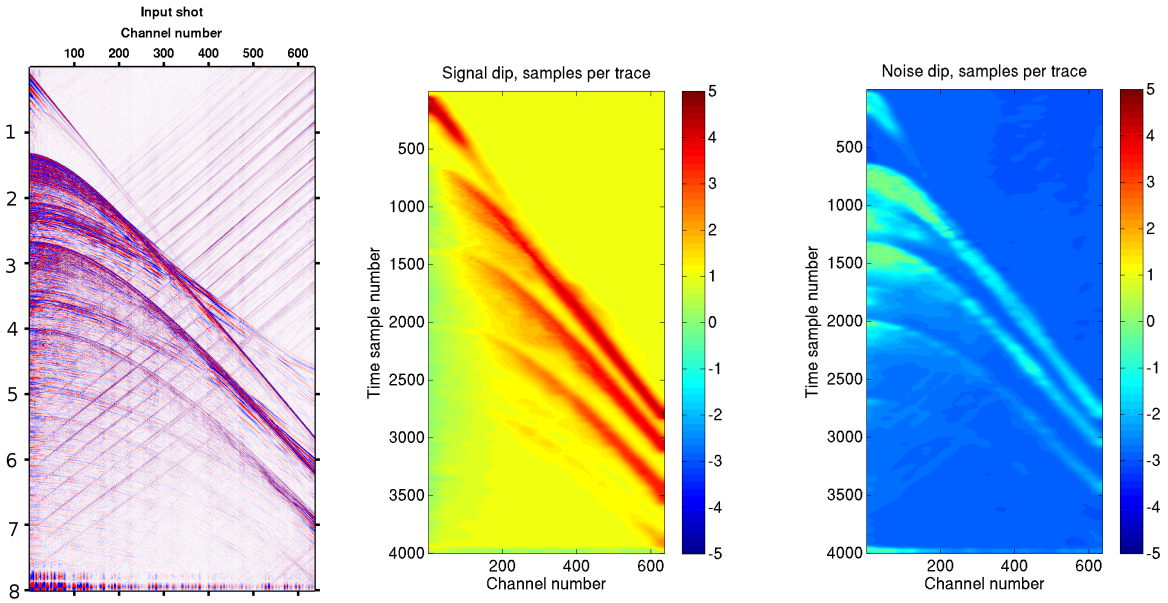


Figure 45: (left) The filtered input shot gather. (middle) The dip field of signal. (right) The dip field of SI noise.

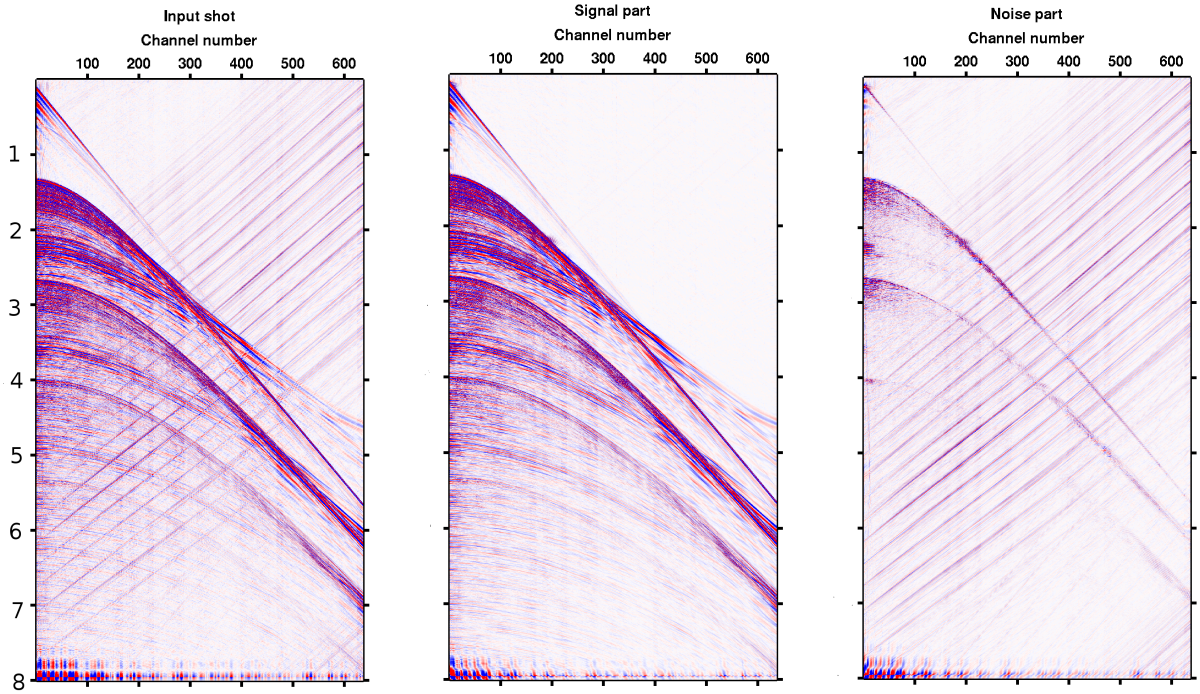


Figure 46: (left) The filtered input shot gather. (middle) The signal part. (right) The noise part. The result of separation after one iteration.

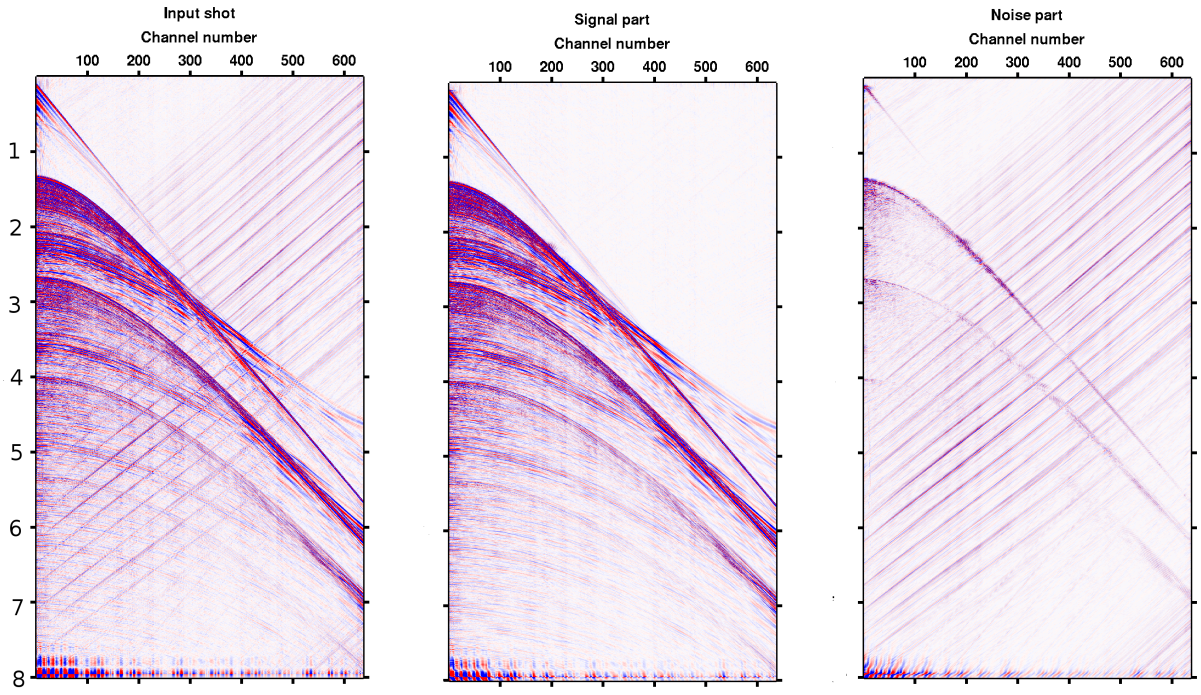


Figure 47: (left) The filtered input shot gather. (middle) The signal part. (right) The noise part. The result of separation after four iterations.

7 CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

The following main conclusions can be drawn from this work.

1. In general, the nonlinear plane wave destruction method is the best choise for SI dip estimation. This is because the estimated dips are reliable (standard deviation values are relatively small even when the method has already failed), it adapts to local changes in dip because the dip operator is small, the method is relatively insensitive to noise $-$ it does not fail even with SNR values as low as 5 or even 2 and, finally, it is fast and suitable to process large amounts of data. The correlation based method is rejected because of being very slow and the structure tensor method because of being too sensitive to noise.
2. The optimal dip filters are those that remove one dip in the T$-$X domain and one corresponding apparent velocity in the F$-$K domain. The optimal finite difference stencil is the one proposed by Claerbout because it has the best dip filter behaviour of all the finite difference stencils. It is also small, thus, able to adapt to local changes in dip. It can safely be applied if dips larger than one sample per trace are not expected. The constrained prediction error filter based stencils are non-aliased stencils able to estimate any dip value. They include only two traces meaning that dipping events are strictly planar. Thus, they are superior to finite difference stencils and have the best dip filter behaviour of all the considered dip filters.
3. SI noise has a minor effect after stacking and post-stack migration because these methods represent effective de-noising techniques. If pre-stack data analyses, like AVO, are to be carried out or pre-stack migration is ought to be done, SI attenuation is still of great importance.

7.2 Future Work

As mentioned in Section 2.3, the phenomenon of SI noise has not yet been fully expained, perhaps because such knowledge is not required for its attenuation. Possible future work could be to investigate further the causes of the SI phenomenon, especially, SI propagation.

SI noise depends on the sea depth, properties of water column (including water layering), seismic source characteristics (such as its size), distance from the SI noise source, properties of the sea bottom (including rock type, thickness of loose sediments) and sea surface (weather conditions) and possibly others. The sea bottom and sea surface properties can be expressed with respective reflection coefficients. A good first step towards an improved understanding is modeling by means of normal modes or ray tracing (Kuperman, 1996). In shallow water areas, such as continental shelf, the sound propagation can be described by waveguide phenomena. The sound propagation can be viewed as constructive and destructive interference of wavefronts reflected at the sea surface and sea bottom (Figure 48).

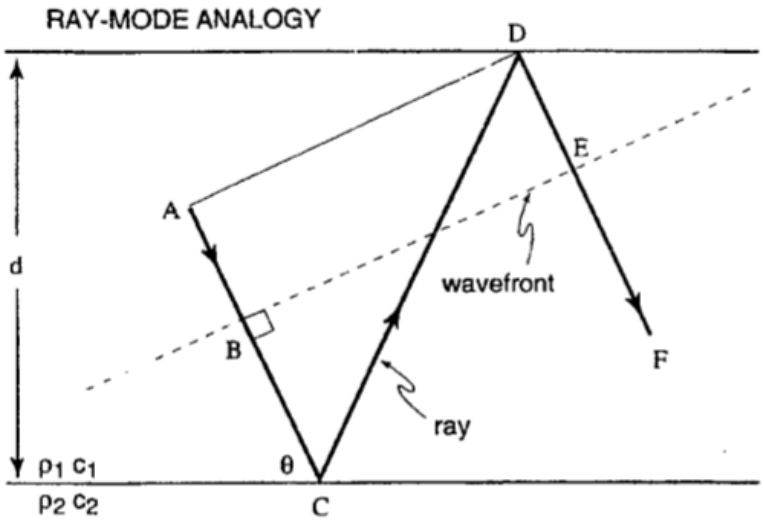


Figure 48: Ray – mode analogy. The ray along path ACDF has a wavefront perpendicualr to the ray. The downgoing rays AC and DF will interfere constructively and form a mode, if points B and E have a phase difference of $n \times 2\pi$, where n is integer number (Kuperman, 1996).

We present a simplified model for SI propagation. The reflection coefficient of the sea surface is set to -1, and the Zoeppritz equations (Zoeppritz, 1919) can be used to estimate the sea bottom reflection coefficient. Figure 49 shows the reflection coefficient for incidence angles between zero and $\pi/2$. The parameters used were estimated from the area of acquisition of the data employed in this work. The water column was assumed uniform, its density was set to 1100 kg/m^3, and its a P-wave velocity equal to 1500 m/s. The underlying solid rock properties were uniform, with a density equal to 1500 kg/m^3, P-wave velocity of 1750 m/s and S-wave velocity of 1029 m/s. This approximate relation between P- and S-wave velocities

was used: $v_P \approx 1.7 v_S$. The real part of reflection coefficient is larger for post-critical reflections than for pre-critical ones and increases with angle. Therefore, it can be assumed that the SI noise travels in the water column. However, here, the Zoeppritz equations are defined for a liquid – solid boundary. Even a thin layer of sediments can change the reflection coefficient (Duncan et al., 2009).

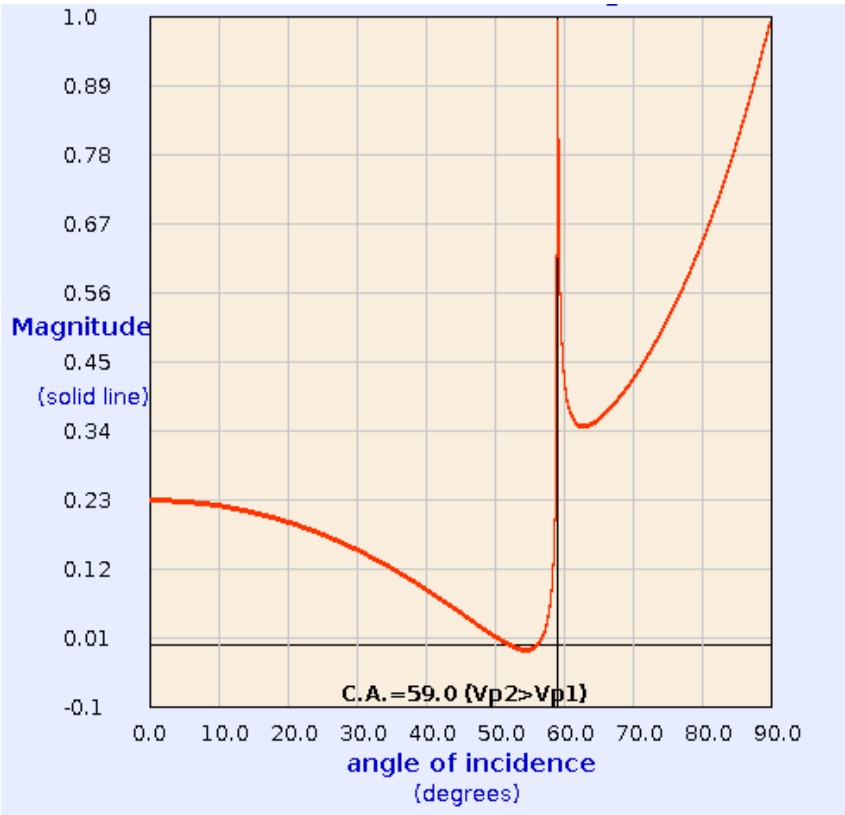


Figure 49: Reflection coefficient for different angles for the estimated parameters (*CREWES Zoeppritz Explorer 2.0* 2005). C.A. stands for 'critical angle'.

As shown in Section 2.3, the SI noise might manifest as a train of "pulses" which perhaps are modes. The time difference between two consecutive peaks is shorter in the begining of the SI noise train and longer toward its end. In the limit of distance from the source it should become constant and equal to the two-way time to the water bottom. The arrival time of the set of pulses that are reflected n times can be predicted by

$$t_n = \sqrt{D^2 + (2nH)^2}, \tag{46}$$

where D is the distance from source to receiver, H is the height of water column (900 m) and n = 1, 2, ... is the number of reflections from the sea bottom. The source and receiver depths were the same for simplicity.

It is interesting that the normal modes model with the parameters representing the area of acquisition of seismic line used in this work gives some features that are observed in the real data: compare figures in Section 2.3, especially, Figure 9, left, with Figure 50.

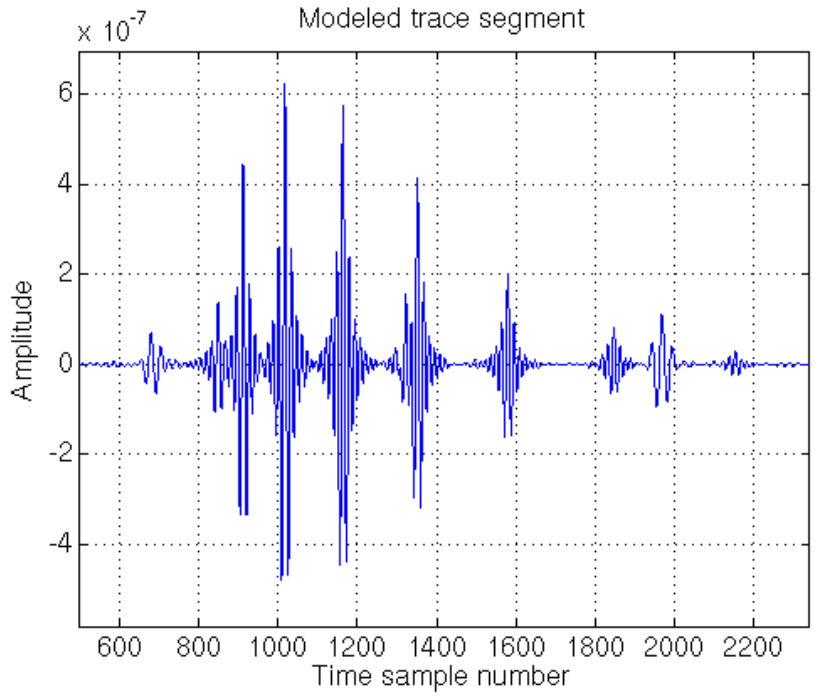


Figure 50: The SI noise modeled with the parameters representing the area of acquisition of seismic data used in this work (see the text). It can be compared with the data examples given in Section 2.3.

Appendix A INVERSE PROBLEMS

Some dip estimation methods, namely, the plane wave destruction methods, are formulated as inverse problems. The linear plane wave destruction method is formulated as a single parameter linear least squares problem. The nonlinear plane wave destruction method is formulated as a global nonlinear least squares problem and is solved by iterative solvers, such as iterative Gauss-Newton method and the conjugate gradient method. Therefore, in this Appendix we present the inverse problem in general as well as relevant methods of solution.

A.1 General Problem Statement

The vector of observations $\mathbf{d}$ and the vector of the model parameters $\mathbf{m}$ are related through the operators which map data space to model space or opposite. If the relation between the data and the model is linear, then the operators are also linear. Such relations can be written as

$$\mathbf{d} = \mathbf{L}\mathbf{m}, \tag{47}$$

$$\mathbf{m} = \mathbf{L}^{-1}\mathbf{d}. \tag{48}$$

Here, $\mathbf{L}$ is the direct, forward or modeling operator (matrix) which models the data using model parameters. It expresses the physical relationships between the measured quantities $\mathbf{d}$ and the parameters of the object of investigation $\mathbf{m}$. The operator $\mathbf{L}^{-1}$ is the inverse or backward operator which maps the data space to the model space. Equations 47 and 48 are called the forward and inverse problems, respectively.

In this Appendix, we focus on the inverse problem where we, based on a set of measurements, want to recover the model parameters. The dimensionality of the data and model vectors can be different. If there are more measurements than model parameters, the problem is overdetermined. If there are more model parameters to find than the measured values, the problem is underdetermined.

If the operator $\mathbf{L}$ is real and orthogonal, then its inverse equals its transpose: $\mathbf{L}^{-1} = \mathbf{L}^{\mathbf{T}}$. If the operator $\mathbf{L}$ is complex and orthogonal, then its inverse equals

its conjugate transpose: $\mathbf{L^{-1}} = \mathbf{L^H}$ (Rowland, 2014). If valid, these relations are very useful because transposing a matrix is much easier than calculating its inverse. The present day seismic inversion is based on approximate inverses – transposes (Claerbout, 2012).

A.2 Least Squares Method

The least squares method finds the best-fitting mathematical or statistical model to the observed data by minimizing the residuals (e.g., the sum of squares of the offsets of the data from the modeled data).

Linear least squares method
In the linear least squares method, the model is linear with respect to model parameters (Weisstein, 2014a; Ruanaidh and Fitzgerald, 1997):

$$\mathbf{y} = b_1\mathbf{x_1} + b_2\mathbf{x_2} + ... + b_M\mathbf{x_M} + \mathbf{e}, \tag{49}$$

and, in the matrix form:

$$\mathbf{y} = \mathbf{Xb}. \tag{50}$$

Here, $\mathbf{y}$, $\mathbf{e}$, $\mathbf{x_m}$ (m = 1, .., M) are N-dimensional vectors, b_m are scalars. In the matrix notation, the vector $\mathbf{e}$ is included in the matrix $\mathbf{X}$. N is the number of observations and thus equations, m varies from 1 to M, where M is the number of parameters. The equation system in Equation 49 is overdetermined because it is required that $N > M$, that is, it has more equations than unknown parameters. The vector $\mathbf{y}$ is called the dependent variable or measured data, vectors $\mathbf{x_m}$ are independent variables or predictor variables, $\mathbf{e}$ is an error term or noise and b_m are model parameters. The matrix $\mathbf{X}$ is of dimensions $N \times M$ and is called a matrix of predictors.

The system in Equation 49 will not provide an exact solution $\mathbf{b}$ which completely describes the observations $\mathbf{y}$, so the closest solution $\hat{\mathbf{b}}$ is sought in the sence of a quadratic norm minimization:

$$Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{Xb}\|^2 = (\mathbf{y} - \mathbf{Xb})^T(\mathbf{y} - \mathbf{Xb}). \tag{51}$$

The function $Q(\mathbf{b})$ is called the objective function, and $\mathbf{y} - \mathbf{Xb} = \mathbf{r}$ is called the residual which we want to minimize. This can be done by setting the derivatives of $Q(\mathbf{b})$ with respect to the parameters in $\mathbf{b}$ to zero. This gives the normal equation system

$$(\mathbf{X^T X})\hat{\mathbf{b}} = \mathbf{X^T y}, \tag{52}$$

which has a unique solution:

$$\hat{\mathbf{b}} = (\mathbf{X^T X})^{-1}\mathbf{X^T y}. \tag{53}$$

From the point of a statistical analysis, Equation 49 expresses a linear regression model with variables satisfying certain conditions. That is, $\mathbf{x_m}$ contains negligible errors (the method only minimizes the error in the dependent variable), the model is linear, all y_i have the same variance and are not correlated with predictors, $\mathbf{X}$ is of rank M. The last condition assures a unique solution $\mathbf{b}$. The solution obtained by (53) is also the most likely solution (Ruanaidh and Fitzgerald, 1997).

Single parameter least squares method

In the simplest case, there is only one fitting parameter b which is then related to the correlation coefficient.

The model $\mathbf{y} = \mathbf{x}b$ with N observations gives an overdetermined system of equations. The minimum of the objective function (sometimes called quadratic function) $Q(b) = (\mathbf{r} \cdot \mathbf{r}) = (\mathbf{y} - b\mathbf{x})^T(\mathbf{y} - b\mathbf{x})$ is found by setting its derivative to zero. The parameter b is equal to

$$b = -\frac{(\mathbf{y} \cdot \mathbf{x})}{(\mathbf{x} \cdot \mathbf{x})} \tag{54}$$

and is related to the correlation coefficient r because the parameter b is the regression coefficient (Weisstein, 2014a)

$$b = \frac{\mathrm{cov}(y, x)}{\mathrm{cov}(x, x)} = \frac{\mathrm{cov}(y, x)}{\mathrm{var}(x)} = \frac{\mathrm{cov}(y, x)}{\sigma_x^2}. \tag{55}$$

The correlation coefficient r describes the reliability of the estimate (strength of dependency between $\mathbf{y}$ and $\mathbf{x}$) and is defined as:

$$r^2 = \frac{[\mathrm{cov}(y, x)]^2}{\sigma_x^2 \sigma_y^2}. \tag{56}$$

In Section 3.3, the slope p is estimated by this single parameter linear least squares method.

A.3 Iterative Methods

Small systems of equations can be solved with direct inverse methods, like Gaussian elimination and back-substitution. Larger systems can be solved iteratively. For dense and not too large systems, the iterative Gauss-Newton method can be applied. For large and sparse systems, the conjugate gradient method can be used in combination with sparse representation.

Iterative methods are used to estimate dips by the nonlinear plane wave destruction method (Section 3.4) and for the dip-based signal and noise separation (Section 5.3).

Nonlinear least squares method and Gauss-Newton iteration
The Gauss-Newton method is an iterative method suitable to solve relatively small systems of equations with dense matrixes. If the objective function is unknown and nonlinear, it can be linearized and the linear least squares method can be applied at each iteration. Even thought the linear least squares method is applied at each step, the method is essentially non-linear (Weisstein, 2014a).

We want to find the minimal value of the residual $r(t)$, where $r(t)$ is in general a vector of N elements. We start by computing its value at some point t_c, then, compute its value at another point $t_c + s_c$. This new point has to be chosen in the way that $r(t_c+s_c) < r(t_c)$. Here, s_c is called a step. This procedure can be linearized as

$$r(t_c + s_c) \approx r(t_c) + J(t_c)s_c, \tag{57}$$

where s_c is the step direction, and $J(t_c)$ is the Jacobian, the matrix of the first derivatives with rows $\nabla r_i(t_c)^T$. To find the stationary point of the function $r(t_c)$, $r(t_c + s_c)$ in Equation (57) is set to approximately zero and solved for the step s_c. This leads to the main least squares equation

$$s_c^{(k+1)} = -\left[J(t_c)^{T(k)} J(t_c)^{(k)}\right]^{-1} J(t_c)^{T(k)} r(t_c)^{(k)}, \tag{58}$$

where (k) denotes the iteration number, $s_c^{(k+1)}$ is the step.

The method requires an initial guess for the step s_c^0 and it is desirable that it is close to the solution, otherwise convergence is not guaranteed. Solving the problem requires the evaluation of the Jacobian and solution for the linear equation system $J(t_c)s_c = -r(t_c)$ (Loan, 2000).

Conjugate gradient method

The conjugate gradient method is an iterative method used to solve sparse systems of linear equations. A good introduction to the conjugate gradient method with visualizations is given in (Shewchuk, 1994). In theory, the method gives an exact solution in N iterations, where N is the number of equations. Systems of equations can be very large and the approximate satisfactory solution can for example be obtained by halting the iterative process after a certain number of iterations or when the change in the solution from one iteration to the next is below a certain threshold.

Suppose, the system of linear equations $\mathbf{Ax} = \mathbf{b}$ needs to be solved. In order to have a solution, the matrix $\mathbf{A}$ should be positive definite (eigenvalues are positive and definite). Two vectors $\mathbf{u}$ and $\mathbf{v}$ are conjugate with respect to the matrix (orthogonal through the matrix $\mathbf{A}$), if $\mathbf{uAv} = 0$. Let P be a set of mutually conjugate directions. Then, the solution $\mathbf{x}_*$ can be expressed as the scaled sum of these directions

$$\mathbf{x}_* = \sum_{i=1}^{N} a_i \mathbf{p}_i, \tag{59}$$

$$\mathbf{b} = \mathbf{A}\mathbf{x}_* = \sum_{i=1}^{N} a_i \mathbf{A}\mathbf{p}_i. \tag{60}$$

For any $\mathbf{p_k} \in P$, $\mathbf{p_i}$ and $\mathbf{p_k}$ are conjugate if $i \neq k$ and then

$$\mathbf{p}_k^T \mathbf{b} = \mathbf{p}_k^T \mathbf{A}\mathbf{x} = \sum_{i=1}^{N} a_i \mathbf{p}_k^T \mathbf{A}\mathbf{p}_i = a_k \mathbf{p}_k^T \mathbf{A}\mathbf{p}_k. \tag{61}$$

From this equation, the scalars a_k can be calculated. It also gives the method for solving the system $\mathbf{Ax} = \mathbf{b}$: we need to find a sequence of N conjugate directions and compute the coefficients a_k.

The quadratic function $0.5\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}$ is iteratively minimized as $\mathbf{x}$ approaches the solution $\mathbf{x}_*$. First, an initial guess $\mathbf{x}_0$ is chosen and the basis vector $\mathbf{p}_0$ is chosen as the derivative (gradient) of the quadratic function at $\mathbf{x} = \mathbf{x}_0$, so $\mathbf{p}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$. Other vectors in the basis are conjugate to the gradient.

Residual at any iteration k is $\mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k$ and it is the negative gradient of the quadratic function at $\mathbf{x} = \mathbf{x}_k$, so the solution moves along the gradient, that is, the direction $\mathbf{r}_k$. The subsequent direction $\mathbf{p}_k$ is a combination of the previous ones

$$\mathbf{p}_k = \mathbf{r}_k - \sum_{i<k} \frac{\mathbf{p}_i^T\mathbf{A}\mathbf{r}_k}{\mathbf{p}_i^T\mathbf{A}\mathbf{p}_i}\mathbf{p}_i \tag{62}$$

and the solution moves to $\mathbf{x}_{k+1} = \mathbf{x}_k + a_k\mathbf{p}_k$, where

$$a_k = \frac{\mathbf{p}_k^T\mathbf{b}}{\mathbf{p}_k^T\mathbf{A}\mathbf{p}_k} = \frac{\mathbf{p}_k^T(\mathbf{r}_{k-1} + \mathbf{A}\mathbf{x}_{k-1})}{\mathbf{p}_k^T\mathbf{A}\mathbf{p}_k} = \frac{\mathbf{p}_k^T\mathbf{r}_{k-1}}{\mathbf{p}_k^T\mathbf{A}\mathbf{p}_k}. \tag{63}$$

A.4 Regularization and Preconditioning

Regularization

In general, the equation system (49) is ill-posed. The opposite of the ill-posed system/problem, the well-posed system satisfies three conditions: a solution exists, the solution is unique, the solution is stable - converges in the iterative solution.

To make the solution exist, unique and the most important, stable, the so called regularization term (constraint) is added to the inverse problem. The second norm of model $\|\mathbf{m}\|^2$ (power), also known as Tikhonov or quadratic regularization, is the most common type of regularization (Neumaier, 1998; Fomel, 2007). Then two terms need to be minimized: the residual and the second norm of model

$$\|\mathbf{d} - \mathbf{L}\mathbf{m}\|^2 + \mu\|\mathbf{m}\|^2 = (\mathbf{d} - \mathbf{L}\mathbf{m})^T(\mathbf{d} - \mathbf{L}\mathbf{m}) + \mu\mathbf{m}^T\mathbf{m}, \tag{64}$$

which is equivalent to require

$$0 \approx \mathbf{L^T}\mathbf{L}\mathbf{m} - \mathbf{L^T}\mathbf{d} + \mu\mathbf{I^T}\mathbf{I}\mathbf{m} \tag{65}$$

and the equation to be solved is

$$\mathbf{m} = (\mathbf{L^T L m} + \mu \mathbf{I})^{-1} \mathbf{L^T d}. \tag{66}$$

The parameter μ is a trade-off parameter. The stability is increased but resolution is decreased (smoothness increased), if μ is large ($\mu \approx 1$) and the opposite, if it is small ($\mu \approx 0.01$).

The regularization operator $\mathbf{R}$ in Equation (66) is equal to the identity matrix $\mathbf{I}$. Other choises exist, such as the derivative of the model $\mathbf{R} = \mathbf{D_1}$ (Fomel, 2002). The regularization term enhances the undesirable features of the model, as in this case, it makes the model rough (Fomel and Claerbout, 2003). If the model space is rough, then the data space is smooth and the opposite. There are also other types of regularizations where other quantities than the second norm are minimized, for example, the first norm of the model. This type of regularization is less sensitive to outliers.

Preconditioning

Preconditioning represents data space regularization (constraint) (Fomel and Claerbout, 2003). In case of the underdetermined system, additional knowledge about the model can be added: $m = Pv$. Here, v is a new model and P is a preconditioning operator which enhances the desirable features of the model, for example, smoothness. If $P \approx R^{-1}$, then

$$\|\mathbf{Lm} - \mathbf{d}\|^2 + \mu \|\mathbf{Rm}\|^2 \tag{67}$$

becomes

$$\|\mathbf{LPv} - \mathbf{d}\|^2 + \mu \|\mathbf{v}\|^2. \tag{68}$$

Preconditioning is superior to regularization for large problems because it leads to faster convergence (Fomel and Claerbout, 2003).

References

Akbulut, R.K., O.K. Sæland, P.A. Farmer, and T. Curtis (1985). "Suppression of seismic interference noise on Gulf of Mexico data". In: *Offshore Technology Conference, 6-9 May 1985, Houston, Texas*.

Bahorich, M.S. and S.L. Farmer (1995). "3-D seismic discontinuity for faults and stratigraphic features". In: *The Leading Edge* **14**, pp. 1053–1058.

Berthelot, A., A.H.S. Solberg, E. Morisbak, and L.J. Gelius (2011). "Salt diapirs without well defined boundaries - a feasibility study of semi-automatic detection". In: *Geophysical Prospecting* **59(4)**, pp. 682–696.

Brittan, J., L. Pidsley, D. Cavalin, A. Ryder, and G. Turner (2008). "A case of SI attenuation in 4D seismic data recorded with permanently installed array". In: *The Leading Edge* **27**, pp. 166–175.

Calvert, R.W., G.J.H. Schoelcher, and H.J. Blankespoor (1984). "Evaluation of a controlled seismic boat interference experiment, Report EP-60410". In: *Shell Internationale Petroleum Maatschappij, The Hague*.

Carter, N. and L. Lines (2001). "Fault imaging using edge detection and coherency measures on Hibernia 3-D seismic data". In: *The Leading Edge* **20**, pp. 64–69.

CGG: BroadSeis (2014). http://www.cgg.com/default.aspx?cid=4894&lang=1. Accessed: 2014-05-10.

Chen, Z., S. Fomel, and W. Lu (2013a). "Accelerated plane wave destruction". In: *GEOPHYSICS* **78(1)**, pp. V1–V9.

— (2013b). "Omnidirectional plane wave destruction". In: *GEOPHYSICS* **78(5)**, pp. V171–V179.

Claerbout, J.F. (1992). *Earth sounding analysis: Procesing versus inversion*. Blackwell Scientific Publications, Inc.

— (2012). *Image estimation by example: Geophysical soundings image construction*. Stanford Exploration Project, http://sepwww.stanford.edu/sep/prof/.

CREWES Zoeppritz Explorer 2.0 (2005). http://www.crewes.org/ResearchLinks/ExplorerPrograms/ZE/ZEcrewes.html. Accessed: 2014-05-20.

Diebold, J. and P. Stoffa (1981). "The traveltime equation, tau-p mapping, and inversion of common midpoint data". In: *GEOPHYSICS* **46(3)**, pp. 238–254.

Duncan, A.J., A. Gavrilov, and L. Fan (2009). “Acoustic propagation over limestone beds”. In: *Proceedings in ACOUSTICS 2009, 23-25 November 2009, Adelaide, Australia*, pp. 1–6.

Elboth, T. and D. Hermansen (2009). “Attenuation of noise in marine seismic data”. In: *SEG Annual Meeting, 25-30 October, Houston, Texas. SEG Technical Program Expanded Abstracts*, pp. 3312–3316.

Elboth, T., I.V. Presterud, and D. Hermansen (2010). “Time-frequency seismic data de-noising”. In: *Geophysical Prospecting* **58**, pp. 441–453.

Fomel, S. (2002). “Application of plane wave destruction filters”. In: *GEOPHYSICS* **60(6)**, pp. 1946–1960.

— (2007). “Shaping regularization in geophysical estimation problems”. In: *GEOPHYSICS* **72(2)**, R29–R36.

Fomel, S. and J.F. Claerbout (2003). “Multidimensional recursive filter preconditioning in geophysical estimation problems”. In: *GEOPHYSICS* **68(2)**, pp. 1–12.

Fookes, G., C. Warner, and R. van Borselen (2003). “Practical interference noise elimination in modern marine data processing”. In: *SEG Annual Meeting, 26-31 October, Dallas, Texas. SEG Technical Program Expanded Abstracts*.

Gersztenkorn, A. and K.J. Marfurt (1999). “Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping”. In: *GEOPHYSICS* **64(5)**, pp. 1468–1479.

Gulunay, N. (2008). “Two different algorithms for seismic interference noise attenuation”. In: *The Leading Edge* **27**, pp. 176–181.

Gulunay, N., M. Magesan, and S. Baldock (2005). “Spatial prediction filters for attenuation of seismic interference noise”. In: *SEG Expanded Abstracts, CR1837*.

Hale, D. (2007). “Local dip filtering with directional Laplacians (CWP-567)”. In: *CWP Project Review Report (Colorado School of Mines, Golden, Colorado, USA, May 21-24)*, pp. 91–103.

Jansen, S. (2013). “Two marine seismic interference attenuation methods”. MA thesis. University of Oslo.

Kommedal, J.H., P.H. Semb, and T. Manning (2007). “A case of SI attenuation in 4D seismic data recorded with permanently installed array”. In: *GEOPHYSICS* **72(3)**, Q11–Q14.

Kuperman, W.A. (1996). “Acoustic propagation modeling in shallow water”. In: *U.S. Navy Journal of Underwater Acoustics* **46(4)**, pp. 275–293.

Landrø, M. and L. Amundsen (2010). “Marine Seismic Sources Part I”. In: *Geo ExPro* **7(1)**. Accessed: 2014-05-10.

Lie, J. E. (1988). “Seismic interference, filtering methods and revised noise limits”. In: *Oceanology ’88: Proceedings of an international conference (Brighton, UK, March 8 - 11)*.

Loan, C.F. van (2000). *Introduction to scientific computing*. Prentice-Hall, Inc.

Luo, Y., W.G. Higgs, and W.S. Kowalik (1996). “Edge detection and stratigraphic analysis using 3D seismic data”. In: *66th Annual International Meeting, SEG, Expanded Abstracts*, pp. 324–327.

Luo, Y., Y.E. Wang, N.M. AlBinHassan, and M.N. Alfaraj (2006). “Computation of dips and azimuths with weighted structural tensor approach”. In: *GEOPHYSICS* **71(5)**, pp. V119–V121.

Marfurt, K.J. (2006). “Robust estimates of 3D reflector dip and azimuth”. In: *GEOPHYSICS* **71(4)**, P29–P40.

Marfurt, K.J., R.L. Kirlin, S.L. Farmer, and M.S. Bahorich (1998). “3-D seismic attributes using a semblance-based coherency algorithm”. In: *GEOPHYSICS* **63(4)**, pp. 1150–1165.

Marfurt, K.J., V. Sudhaker, A. Gersztenkorn, K.D. Crawford, and S.E. Nissen (1999). “Coherency calculations in the presence of structural dip”. In: *GEOPHYSICS* **64(1)**, pp. 104–111.

MathWorks Inc. Documentation center (2014). *Cross-correlation*. http://www.mathworks.se/help/signal/ref/xcorr.html. Accessed: 2014-03-02.

Morelatto, R. and R. Biloti (2013). “Structure enhancing filtering with the structure tensor”. In: *Thirteen International Congress of The Brazilian Geophysical Society*.

Neidell, N.S. and M.T. Taner (1971). “Semblance and other coherency measures for multichannel data”. In: *GEOPHYSICS* **36(3)**, pp. 482–497.

Neumaier, A. (1998). “Solving ill-conditioned and singular linear systems: a tutorial on regularization”. In: *SIAM Review* **40**. Accessed: 2014-05-12, pp. 636–666.

Proakis, J.G. and D.G. Manolakis (1996). *Digital signal processing: principles, algorithms and applications*. Prentice-Hall, Inc.

Randen, T., E. Monsen, C. Signer, A. Abrahamsen, J.O. Hansen, T. Sæter, and J. Schlaf (2000). “Three-dimensional texture attributes for seismic data analysis”. In: *SEG Technical Program Expanded Abstracts* **19**, pp. 668–671.

Rowland, T. (2014). *Unitary Matrix*. http://mathworld.wolfram.com/UnitaryMatrix.html. Accessed: 2014-05-10.

Ruanaidh, J.J.K. and W.J. Fitzgerald (1997). *Numerical Bayesian methods applied to signal processing*. Springer.

Scales, J.A. and R. Snieder (1998). "What is noise?" In: *GEOPHYSICS* **63(4)**, pp. 1122–1124.

Schleicher, J., J.C. Costa, L.T. Santos, A. Novais, and M. Tygel (2009). "On the estimation of local slopes". In: *GEOPHYSICS* **74(4)**, P25–P33.

Sercel: Air gun for seismic survey (2014). http://www.nauticexpo.com/prod/sercel/air-guns-seismic-surveys-40158-326494.html. Accessed: 2014-05-10.

Shewchuk, J.R. (1994). *An introduction to the conjugate gradient method without the agonizing pain*. http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf. Accessed: 2014-03-26.

Sobel, I. and G. Feldman (1968). *A 3x3 isotropic gradient operator for image processing*. Presented at a talk at the Stanford Artificial Project.

Subsea World News (2013). http://subseaworldnews.com/2013/06/12/cgg-to-acquire-broadseis-3d-survey-offshore-colombia/. Accessed: 2014-05-10.

Szeliski, R. (2010). *Computer vision: algorithms and applications*. http://szeliski.org/Book/. Accessed: 2014-04-09.

Teigen, O., K. Ozdemir, B.A. Kjellesvig, N. Goujon, and Pabon J. (2012). "Characterization of noise modes in multicomponent (4C) towed-streamers". In: *74-th EAGE Conference & Exhibition incorporating SPE EUROPEC 2012, Copenhagen, Denmark, 4-7 June 2012*.

Thiran, J.P. (1971). "Recursive digital filters with maximally flat group delay". In: *IEEE Transactions on Circuit Theory* **18**, pp. 659–664.

Vliet, L.J. van and P.W. Verbeek (1995). "Estimators for orientation and anisotropy in digitized images". In: *ASCI'95, Proc. First Annual Conference of the Advanced School for Computing and Imaging (Heijen, NL, May 16-18), ASCI*, pp. 442–450.

Warner, C., J. Brittan, R. van Borselen, and G. Fookes (2004). "Attenuation of interference noise on marine seismic data". In: *EAGE 66th Conference & Exhibition - Paris, France, 7-10 June 2004*.

Weisstein, E.W. (2014a). *Least Squares Fitting*. http://mathworld.wolfram.com/LeastSquaresFitting.html. Accessed: 2014-05-02.

Weisstein, E.W. (2014b). *Radon Transform*. http://mathworld.wolfram.com/. Accessed: 2014-03-28.

Wenz, G.M. (1962). “Acoustic ambient noise in the ocean: spectra and source”. In: *The Journal of the Acoustical Society of America* **34(12)**, pp. 1936–1956.

Zoeppritz, K. (1919). “Erdbebenwellen VII. VIIb. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen”. In: *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse* **1**, pp. 66–84.

Årre, V. (2013). *Consistent dip estimation for seismic imaging*. Patent No.: US008463551B2.